On some aspects of quantitative risk management:

theoretical and empirical implications for agricultural goods

Meng (Simon) WANG

Department of Mathematical Sciences

Supervisor: Dr. Hirbod Assa, Dr. Corina Constantinescu

A thesis submitted for the degree of

Doctor of Philosophy

University of Liverpool 2018

This thesis has 160 pages, 145 of them are marked by Arabic numbers.

This thesis is dedicated to
My parents and my wife.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Dr. Hirbod Assa for the continuous supports of my Ph.D study and related research, for his patience, motivation, and immense knowledge. It has been an exciting and rewarding 4-years' to my life. His guidance helped me in researching and writing of this thesis as well. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank Dr. Corina Constantinescu, Dr. David Siska and the rest of IFAM team for their insightful comments and encouragements, but also for the opportunities and helps which have incented me to widen my research from various perspectives.

I would like to thank Dr Şule Şahin and Dr. Keivan Mallahi-Karai for their advice and comments on the thesis.

My sincere thanks also goes to the Institut de la finance structurée et des instruments dérivés de Montréal, the Society of Actuaries of North American, RARE program, Business Fateway of the University of Liverpool, Department of Mathematical Science, AOX Captial Limited and AlgoLib LTD for their generous funding in supporting my research and providing me greater opportunities.

I thank the University of Liverpool and the School of Management for providing me with the platform to apply my research to practice and thank the Liverpool Doctoral College for all the knowledge transfer programmes provided to me.

Last but not the least, I would like to thank my family: my parents and my wife for supporting me throughout writing this thesis and my life in general.

Abstract

Thesis title: On some aspects of quantitative risk management: theoretical and empirical implications for agricultural goods

Author: Meng (Simon) WANG

To the best of our knowledge there are only a few papers on the area of quantitative risk management that is applied to agricultural markets. This thesis aims to fill this gap by studying different aspects of the agricultural data sets within three main chapters.

In the first part (Chapter 2), we focus our attention to agricultural insurances that hedge against the losses from freezing temperature and frost. Our approach is to define a model of harvest losses based on temperature and as a result design and price different type of frost insurances. The results of this part are examined on the real data from San Joaquin in California, US.

In the second part (Chapter 3), we have studied the Hurst exponent, by introducing a new method to estimate it. Then we applied the new method to the commodity data set. The results show that commodities can be differentiated by their Hurst value to agricultural and non-agricultural goods. This shows the agricultural goods have different time series characteristics (i.e., in terms of long memory) and the way they need to be hedged are different than non-agricultural goods.

In the third part (Chapter 4), we have studied the insurance markets on the UK agricultural price indexes. We look at the two sides of the market, demand side and supply side, and see if such an insurance market can be run in the UK. We assumed that the demand side (the farmers) are simply risk minimizers and the supply side (investors and insurance companies) are profit maximizers. First we design an optimal contract from the farmers point of view, and then we show how a portfolio of optimal insurance contracts can benefit insurance companies. The data of the UK agricultural index prices show that returns from this market is comparable to FTSE portfolios.

In the last chapter we provide some new insights for further research in the future.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This is a thesis on some aspects of risks management applied to agricultural markets. We shall see how applied mathematics helps with risk analysis, as well as the delivery of several models that can be applied in the industry for practical usages. Three main chapters on slightly different topics will be presented in this work, and we shall provide our humble understanding of these fields.

Before we start with the main chapters, this chapter will be dedicated to provide the necessary backgrounds and terminologies, so the readers may have the same understanding when different terms are used and referenced in the thesis.

When we talk about how applied mathematics can help with risk management in insurance, essentially, we are talking about the creation, exploration and use of the data sets, algorithms, models and data analyses.

More specifically, in almost all real problem settings, one usually starts with a data set. Attention needs to be paid to the frequency, volatility, stationary and memoryness of the individual data sets as well as parallel time series that provide information and define the fundamental problems [25, 33]. The characteristics of these data act as filters in the selection and creation of potential algorithms and models. For example, in cases where long-memory processes present, using a traditional Auto-Regressive-Moving-Average (ARMA) model will not be adequate; instead, Long-Range Dependence (LRD) models should be employed (Chapter 3).

Theoretical results usually require assumptions to work and to make them useful in the real world. Thus, the modelling of risks is yet another issue that can harm our attempt to introduce the correct models. This is why we need analysis - the bridge between a real problem that need to be solved, and the theoretical mathematical

solutions that have been developed behind it. Analyses explain the purpose of the current study, demonstrates how well the models work, and what still needs to be done.

In the following, the structure of each chapter follows the same process as above: from the real data to a proposed model, and from mathematics to real applications. We shall see in Chapter 2, a method is proposed and tested to price insurance contracts when claims or loss data are not directly available. In Chapter 3, we will discuss how to introduce a recursive parameter estimation method on long memory process. Lastly, in Chapter 4, we discuss how to price and find the optimal contracts for an insurance contract with regarding the demand and supply side. All the chapters are based on real-data calibrations and the results are directly applicable to risk management businesses.

1.1 Definition and properties for risks

Risk is a complicated concept that can be interpreted in different ways. In this thesis, we are more interested in measuring (or assessing) the (negative) impact of uncertainties about an event; this is 'risk'. While the uncertainties about an event is objective and is only concerned with the modelling, risk assessment can be totally subjective and depends on the agents' risk aversion level. In this thesis, the risks that we are going to talk about are mainly financial risks, i.e. the profits or losses that are quantifiable along with financial contracts. Before we provide any mathematical definition of risk, we provide the following example for a clearer understanding of this concept.

Example 1. Assume for Cattle price index. One wants to issue a financial contract on it. The contract payoff is calculated at maturity for the difference between market price (S) and settlement price (K). Figure 1.1.1 and Table 1.1.1 demonstrate the historical performances for Cattle price index and its statistical properties. As one can see, the empirical data that is very volatile and showing no easily recognisable pattern.

Figure 1.1.1: Historical Deadweight Cattle performance

Assume that the financial product is a rolling contract that settles on the monthly price differences and the contract holder will gain from price rises and suffer from price drops. If the trading is happening every month for over the past 138 months (note, the last month's contract cannot be closed without new market data), we denote the

	Deadweight Cattle Price	Cattle Price Difference
Count	139.00	138
Mean	302.00	1.27
Std	58.68	6.93
min	192.10	-18.88
25%	268.70	-2.75
50%	319.30	0.96
75%	349.81	5.90
max	400.80	27.45

Table 1.1.1: Statistical summary for Deadweight Cattle data

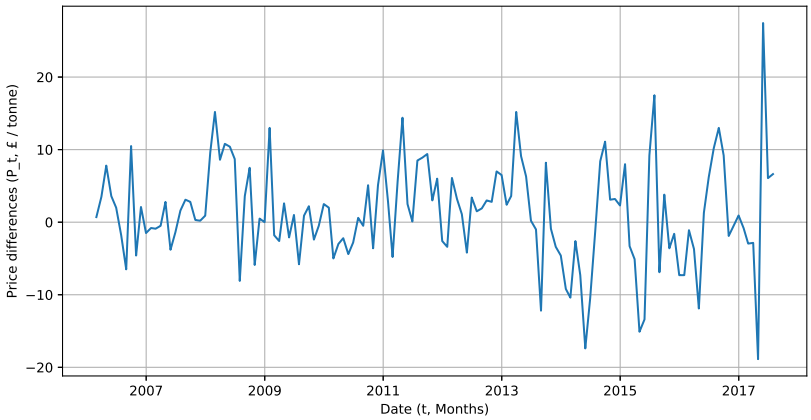


Figure 1.1.2: Historical Deadweight Cattle monthly price differences

trader's payoff by P_t, and clearly $P_t = S_t - K_t = S_t - S_{t-1}$, where t is the months and takes integers from 2 to 139. Figure 1.1.2 shows the historical payoffs for the same Cattle data. As one can see, the shape and pattern of the payoff P_t are significantly different from the index price data S_t. Table 1.1.1 provides a more detailed view on the trader's historical positions. The mean value and 50% percentile in Table 1.1.1 show that the trader is making profits for most of the time, and, in the long run, the huge volatility indicates the large uncertainties are involved in the trading of this product.

Thus, we use the word 'risks' in financial frameworks to indicate the potential of gains or losses in positions (note, risks refer not only to losses, but also applies to profits).

Furthermore, we can simply look at different quantities as a way to measure the risk of the difference in the prices (see Table 1.1.2). Actually different quantities can

represent different risk aversion attitudes towards risk.

Percentiles	90.5	92.5	95.5	96.5	97.5	99.5
VaR values	9.60	10.37	13.00	13.74	15.16	25.36

Table 1.1.2: Quantities for monthly Cattle price losses

In summary, if one could build a probability measure on scenarios, we could focus on the resulting distribution of S and try to measure the risk in terms of moments or quantiles from a probabilistic model. However, a classical measure of risk such as the variance does not usually capture the basic asymmetry in the financial interpretation of S where it could be only the downside or upside that matters. Thus, one will need measures like Value at Risk (VaR) to take into account risk on one side tail of the distribution.

This thesis will be focus on risks with financial senses, there are different types risks in this area namely: asset-backed risk, credit risk, foreign exchange risk, liquidity risks, market risk, operational risk, model risk and etc. In the following sections, when we refer to 'risk', we refer to the potential quantifiable events that may lead the contract holder or issuer or both to a bad position or to a scenario that can cause financial losses.

As indicated in the example, due to the underlying market uncertainty, for any position holder, it is natural to ask the following questions.

Question 1. *How to measure the financial risks (particular the downside ones) for some risk exposure?*

(Note that, profit and loss are not good indicators for risk, e.g., the same £18.88 loss in Cattle price or Apple's stock means two different things.)

Question 2. *If a financial downside risk is anticipated, what are the tools one can use to control losses?*

(This is important due to the uncertainty of the market or unfortunate events.)

Question 3. *If a certain financial risk is identified, how can one price the protection cost for such a risk?*

(i.e., what is the cost of the protection against potential risks, and what types of protection can I expect?)

Question 4. *Can financial risks being forecast if one can simulate events that have long memories?*

(For instance, in the previous example uses historical data, what risk event will happen beyond historical data?)

Question 5. *Can financial risks be reduced by studying the supply and demand of the underlying markets?*

(i.e., when designing financial contracts, can we put demand and supply into considerations and try to minimise risks beforehand?)

We shall provide our answers to these questions in the following chapters. The rest of this chapter provides a review of some of the necessary concepts that we will use later, and we try to answer Question 1 and 2. Section 1.2 discusses the risk measures, their mathematical definitions and their implications. Section 1.3 introduces the formal definitions of insurance and reinsurance contracts, different protection types and traditional pricing methods. Section 1.4 presents several modelling techniques as well as a short analysis of different models and parameter estimations. All these sections come with examples for clearer interpretations. Section 1.5 summarises the data sets we used in the thesis. Lastly, we list our contributions and further potential developments in section 1.6. Following this chapter, Question 2 and 3 will be discussed using a financial mathematics method developed in pricing frost insurances in Chapter 2. In Chapter 3, we focus on answering Question 4 by introducing a new parameter estimation method, where a more general time series model is used to model long-range dependency. Question 3 and 5 are studied again in Chapter 4, where a more general framework on minimising the risk is proposed, and the prices of contracts are drawn from market data.

1.2 Risk measures

Following [35], we consider a probability space $(\Omega, \mathcal{F}, P)$. Here Ω is the set of scenarios, $\mathcal{F}$ is a σ-field of the subsets of Ω, called the set of events and P is a probability measure. We denote the set of all random variables on Ω by $L^0(\Omega) = L^0$. For any $X \in L^0$ the cumulative distribution function (CDF) and probability density function (PDF) of X are denoted by F_X and f_X, respectively.

To answer Question 1, we go back to the first example to see what the statistical characteristics of risk looks like and how we can measure it effectively. In order to make different underlying assets directly comparable for risk management, profit or loss (price differences) rather than price itself are usually used. Also, to provide better modelling results and to meet model assumptions, we will use differences $(S_t - S_{t-1})$, returns $(\frac{S_t - S_{t-1}}{S_{t-1}})$ and log returns $(\ln S_t - \ln S_{t-1} = ln(\frac{S_t}{S_{t-1}}))$, see [33] for details. Figure 1.2.1 shows the histogram for Cattle data in Example 1 when four different measurements (price itself, price difference, return and long returns) are applied to the price data.

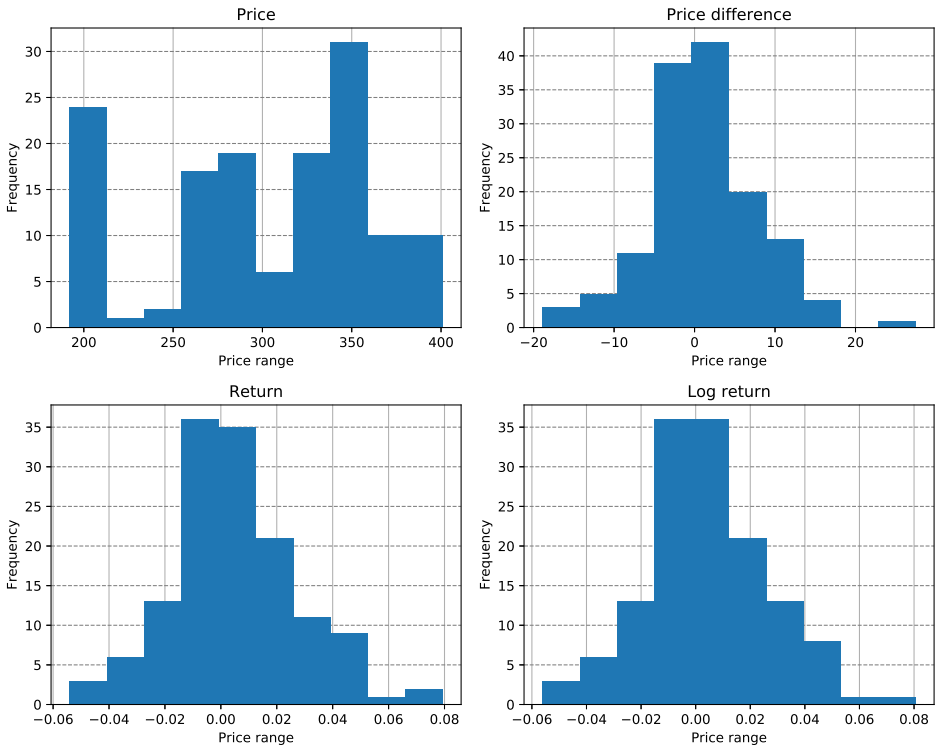


Figure 1.2.1: Cattle price histograms

It is straightforward to observe the risk exposures from the histograms, the percentage of losses and the probability of the events. However, comparing the histograms may not necessarily be a good idea, particularly in cases where only negative risks or the tail events (due to insufficient data) are of interest. Thus, a single number that can summarise the risk can be very useful in a risk management framework. Straightforward choices are different moments from the sample data or the empirical densities, i.e. mean, variances, kurtosis, skewness etc. Beside moments, to deal with the asymmetry, some common risk measures including VaR (value-at-risk), CVaR(expected-shortfall, conditional value-at-risk) are also useful. However, given different purposes, a more general definition for the measures is summarised below from [2, 7] and [71].

Definition 1. Let us consider $\mathcal{D} \subset L^0$ consists of all loss variables. A risk measure ρ (premium) is defined as a non-decreasing mapping from $\mathcal{D}$ to the real numbers $\mathbb{R}$ that $\rho(0) = 0$.

A risk measure can have one or few properties from the following list

1. Positive homogeneity of degree one i.e., $\rho(\lambda X) = \lambda\rho(X), \forall\lambda \geq 0, X \in L^0$.
2. Cash in-variance i.e., $\rho(X + c) = \rho(X) + c\rho(1)$.
3. Monotonicity i.e., $\rho(X) \leq \rho(Y)$, for any two $X \leq Y$ in $\mathcal{D}$;
4. Law invariance i.e., $\rho(X) = \rho(Y)$ if $F_X = F_Y$.
5. Co-monotone additivity i.e., $\rho(X + Y) = \rho(X) + \rho(Y)$, for any two non-decreasing real functions f, g and a random variable Z such that $X = f(Z)$ and $Y = g(Z)$.
6. Continuity from above i.e., if $X_n \downarrow X$ a.s., then $\rho(X_n) \downarrow \rho(X)$.
7. Continuity in probability if $X_n \to X$ in probability, then $\rho(X_n) \to \rho(X)$.
8. Continuity in L^p i.e., if $X_n \to X$ in L^p, then $\rho(X_n) \to \rho(X)$.

In many applications, we need to put more emphasis on particular events; for instance, we need to assign larger weights to larger losses (we need to be more pessimistic). In the following we introduce Distortion Risk Measures from [32] and[110]. Distortion Risk Measures adjusts risk values by reassigning weights to the density of a loss variable.

Definition 2. A distortion function g is a non-decreasing function from $[0,1]$ to $[0,1]$ such that $g(0) = 0$, $g(1) = 1$, and a distortion risk measure (premium) associated with g is defined as follows

$$\pi_g(X) = \int_0^\infty g(\bar{F}_X(x))dx, \tag{1.2.1}$$

where $\bar{F}_X(x) = 1 - F_X(x)$ is the survival function of $X \in L^0$.

For example, Value at Risk (VaR_α) is a popular example of distortion risk measure with a distortion function

$$g(x) = \begin{cases} 0 & 0 \leq x < 1-\alpha \\ 1 & 1-\alpha \leq x \leq 1 \end{cases},$$

where $\alpha \in (0,1)$ is the confidence level. Therefore, as shown in Figure 1.2.2, VaR is defined as

$$\begin{aligned} \text{VaR}_\alpha(X) &= \inf\{u \in \mathbb{R} : \mathbb{P}(X > u) \leq 1-\alpha\} \\ &= \inf\{u \in \mathbb{R} : F_X(u) \geq \alpha\}. \end{aligned} \tag{1.2.2}$$

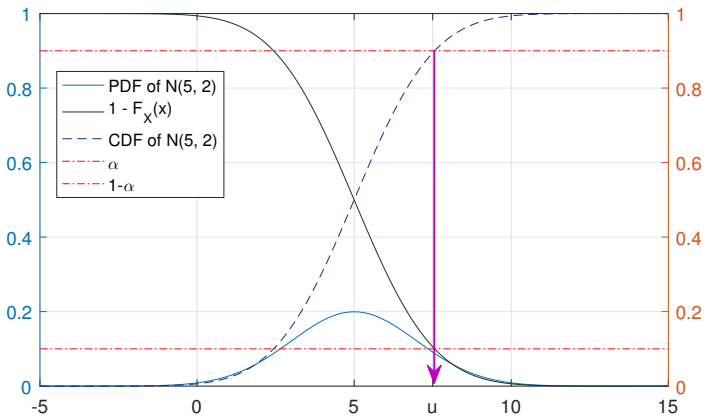


Figure 1.2.2: VaR and distortion measures in a premium sense, when $\alpha = 0.1$ and $\mathcal{N}(5,2)$ is followed.

Conditional Value at Risk (CVaR), whose distortion function is given by

$$g_{\mathrm{CVaR}_\alpha}(t) = \frac{t}{1-\alpha} 1_{[0,1-\alpha]}(t) + 1_{[1-\alpha,1]}(t),$$

is another popular risk premium that can also be represented in terms of VaR as follows

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_t(X) dt. \tag{1.2.3}$$

Finally, the Wang premium function (see [110] and [111]) is introduced by the following distortion function known as Wang's transformation

$$g_\beta(x) = \Phi(\Phi^{-1}(x) + \beta), \tag{1.2.4}$$

where $\beta \in \mathbb{R}$ is a real number and Φ is the CDF of the standard normal distribution with mean equal to zero and standard deviation equal to one. The parameter β is proportional to the market price of risk, meaning that it measures a reward paid for holding a risky position. That is why the larger the β, the larger the market price of risk will be.

1.3 Insurance and reinsurance

Once risk is directly measurable and can be quantified by using a risk measure, Question 2 becomes much easier to answer. As a result, we can think of transferring risk to another party by introducing insurance and reinsurance contracts. In this section we briefly discuss insurances and reinsurances.

Numerous tools exist for risk protection and hedging, see [92], [6] and [84] for details. For example, one can construct a portfolio in the market that has negative correlations with the risk exposure and hedges the risk purely using financial mathematics methods, or one can introduce derivatives and insurances to transfer the risk (like call and put options). While derivatives are not (or cannot be) available for all types of risks, insurances can always be considered as a common risk protection tool in the real world. For example, for agriculture goods, the main product traded in the current derivative markets in the UK and Europe are Wheat and Rapeseed oil futures where these derivative markets need very high entrancing barriers in terms of required funds and follow a very strict regulation. It is almost impossible for regular farmers to protect their losses in these markets for their own goods. Thus, insurances become a necessary risk protection tool to protect against financial losses.

Following [31], insurance here is a form of risk management primarily used to hedge against the losses. Recall that, in the previous example, we have a market price denoted as S_t, a settlement price K_t. Assume instead of trading on the Cattle price market, a Cattle farmer wants to hedge against price changes with an insurance policy. That is one example. The other example can be damages to the real harvest where we introduce insurance on the yield denoted by Y_t. In general, we denote the loss variable with X_t to denote the loss from prices, indexes, yield or any other variable related to agricultural goods.

Definition 3. Let f_{in} denotes an insurance policy writes on some risk variable X_t at time t, the policy also has a settlement value K. $t \in [0,\ T]$ is the index for time, and T is the contract length. An insurance contract to protect against falls (prices, yield, revenue etc.) promises to pay the policy holder the amount of $f_{in}(X_t) = \Sigma_0^T max(0,\ K - X_t)$ at the end or any time during the contract period $[0,\ T]$ for a protect fee of p. We usually call p the premium of the insurance.

For the same reason, those who issue insurances, called underwriters or insurance companies, also need protections for their risk exposures; thus, we call these types of insurance contracts reinsurance. Reinsurance is insurance that is purchased by an insurance company from other insurance companies directly or through a broker as a risk management tool, see [26] for details.

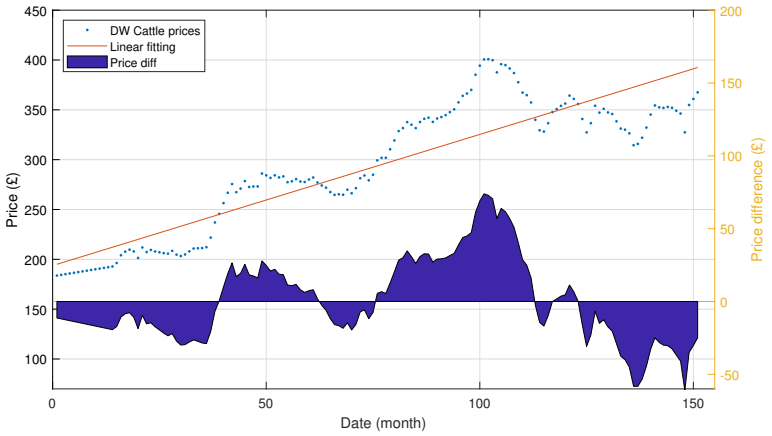


Figure 1.3.1: Demonstration on insurable risks

Example 2. For instance, following Example 1, let us assume an underwrite writes a monthly insurance on Cattle prices based on our indices. The settlement price is monthly adjusted with a linear forecast model as the red line shows in Figure 1.1.2. The insurer agrees to pay the policy holder the price difference in the case that there is a price drop every month. The negative blue area plot indicates the insurer's risks, thus the policy holder's claims. If the contract is not a monthly contract but a yearly, then, the settlements will be the accumulated values for the 12 months. Such an insurance contract protects the policyholder's risks for price falls.

In other cases, if the underwriter or the policyholder needs no full protection for all risk exposures, three difference policies are created: Quota-share insurance, Stop-loss insurance and Excess-of-loss insurance. Note, the same names also apply to reinsurance contracts, and, in this section, we shall only give definitions for protections over drops in value (e.g., following Example 1 for clarity and see [27] for details of definitions).

Definition 4. A Quota-share reinsurance is a type of insurance where only a certain percentage of claims will be covered by the underwriter. Denote f_{quota} as the final total settlement of a Quote-share insurance for the protection of price drops, then

$$f_{quota}(K,\, t) = \sum_{t=0}^{T} \alpha\, max(0,\ K - X_t)$$

where $\alpha \in \mathbb{R}$ is the Quota-share ratio, K is the settlement value, X_t is the loss variable at $t \in [0, T]$ for maturity $T \in \mathbb{R}$.

Definition 5. Stop-loss reinsurance is where the accumulated settlement is only paid from a certain level to a certain level; such a contract is designed to eliminate the risk of huge accumulated risks. Denote f_{stop} as the final total settlement of a Stop-loss policy protecting values drops, then

$$f_{stop}(K,\, t) = \begin{cases} 0, & \sum_0^T (K - X_t) < B_L \\ \sum_{t=0}^T (K - X_t), & B_L \leqslant \sum_{t=0}^T (K - X_t) \leqslant B_U \\ \sum_{t=0}^T (K - X_t), & \sum_{t=0}^T (K - X_t) \geqslant B_U \end{cases}$$

where $B_L,\, B_U \in \mathbb{R}$ and $B_U > B_L \geqslant 0$ are the Stop-loss barriers or retention levels, K is the settlement value, X_t is the loss variable at $t \in [0, T]$ for maturity $T \in \mathbb{R}$.

Definition 6. Excess-of-loss reinsurance is just when stop-loss insurance applies to individual claims rather than the accumulated ones. Define f_{xl} as the cumulative settlement of an Excess-of-loss insurance at the end of the contract to hedge against price drops, then

$$f_{xl}(K,\, t) = \sum_{t=0}^{T} \left(max(0,\ C_U - X_t) - max(0,\ C_L - X_t)\right)$$

where $C_L,\, C_U \in \mathbb{R}$ and $C_U > C_L \geqslant 0$ are retention levels for individual claims, K is the settlement value, X_t is the loss variable at $t \in [0, T]$ for maturity $T \in \mathbb{R}$.

Note, if the insure is only taking one claim for the whole period of a contract, then the Stop-loss policy is equivalent to the Excess-of-loss policy.

Before, we start the data modelling, we provide an initial attempt to answer Question 3, by pricing insurance contracts and finding the optimal retention levels. An insurance contact has to reflect the fair premium price of a risk transformation tool. Traditionally, the price of an insurance contract is based on the amount of the risks that the insurer is going to take. If no historical claim data are available, a careful study of the underlying market is necessary. In Chapters 2 and 4, two different pricing methods are proposed, one by considering the basis risk embedded in the frost events and one based on market price indexes respectively.

For the completeness of the thesis, we provide a common practice on insurance pricing here. In the industry, in general the price of an insurance contract is given by a risk premium principle of the potential claims. This value can be adjusted by another loss factor regarding the catastrophic event evaluation and safety loading. The expected premium rule is used in our example here; the expected premium rule is simply the expected values of potential claims. We also assume there is no adjustment needed for insurance prices due to rare or catastrophic events.

Recall the price differences in Figure 1.1.2, for a monthly insurance that protects price falls and can only be exercised once. We denote claims by P_t. Following the previous notations, one can have $P_t = max(0,\ K_t - S_t)$ where K_t is the linear fitting values and S_t is the underlying asset prices. Define the negative claims ratio as:

$$r_P(t) = \frac{P_t}{S_{t-1}} = \frac{max(0,\ K_t - S_t)}{S_{t-1}}.$$

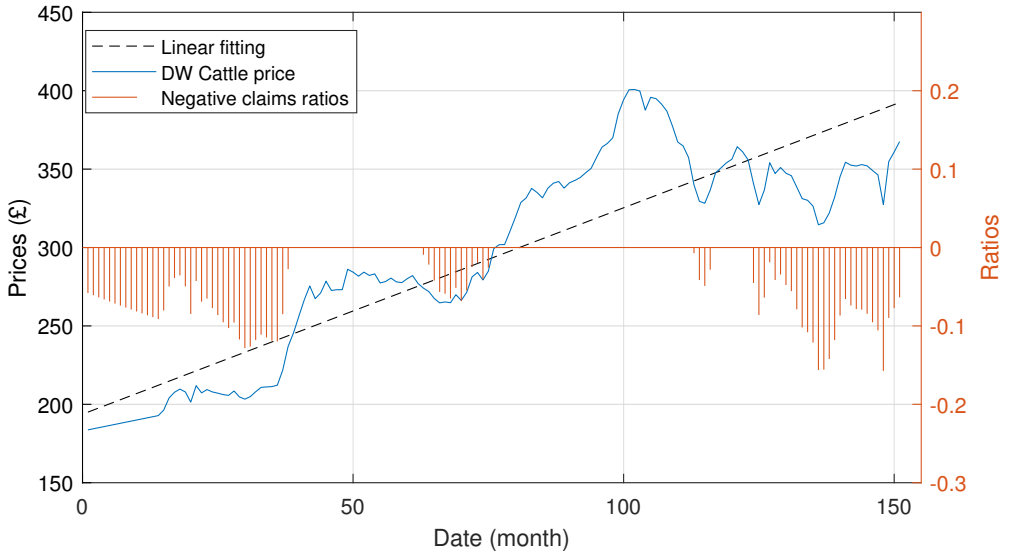


Figure 1.3.2: Negative claims ratios

Figure 1.3.2 shows the negative claims ratios in stem lines. Then the price of the insurance policy following expected premium rule is defined as

$$p = \sum_{t=0}^{T} \frac{r_p(t)}{T}$$

In this case, $p = 7.37\%$ for the plain insurance contract with no catastrophic event nor safety loading added.

1.4 Modelling and forecasting

Recall Figure 1.2.1, where we plot the histograms for different price changes, one can see, the density of the risk exposures provides us almost all the information one can observer from the data. However, modelling is an inevitable practice, since the data on its own cannot answer all the questions we have. For instance, any risk measure that we introduce summarises the density into a single number by taking a moment, or looking at a specific area, or changing the shape of the density. This is in principle impossible without fitting a good model to the real data. In this thesis, we talk about Question 4 for three reasons: first of all, modelling data can provide us with more mathematical flexibility to capture and mimic the data behaviours that are not within the empirical data; second, for problems with very limited raw data sets, modelling can solve the problem to certain degree that most of the common risk management tools will become useful, third we can use calibrated model to forecast future events and distinguish statistical or even economical properties between different products.

We shall give a short introduction of modelling, looking at the statistical properties that we need in this thesis. Most of the models we use fall into the category of stochastic models. To be consistent with the literature, we separate them into stochastic and time series models.

1.4.1 Stochastic models

Stochastic processes have been studied from the late 19th century for better understanding financial markets and the Brownian motions. [105] firstly described the mathematics behind a Brownian motion with the method of least square. Then, Bachellier (see [9]) presented a stochastic analysis of the stock markets in his PhD thesis. Einstein [34] subsequently presented a way to indirectly demonstrate the existence of molecules, describing the fundamentals of a stochastic model. As we discussed earlier, stochastic models have been developed in the last few decades for pricing financial derivatives, and they are an essential part of modelling financial markets.

A stochastic process $W(t)$ is called a standard Brownian motion (also called a Wiener process) if it satisfies:

1. $W(0) = 0$, almost surely;

2. $W(t)$ has independent increments;

3. $W(t) - W(s) \sim N(0, t-s)$, $0 \leq s \leq t$.

4. $W(t)$ is pathwise continuous i.e., for any $\omega \in \Omega$, $t \mapsto W(t, \omega)$ is continuous in t.

One of the fundamental models we shall use later is the Geometric Brownian Motion (GBM). GBM is used to model stock prices in the Black-Scholes model and is the most widely used underlying model to capture stock price behaviour (see, [49]). To be specific, a GBM is a continuous-time stochastic process in which the logarithm of the randomly varying quantity follows a Brownian motion with drift. A stochastic process S_t is said to follow a GBM if it satisfies the following stochastic differential equation:

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

where W_t is a Brownian motion, drift μ and volatility σ are constants.

Following the GBM, the second contentious time stochastic model is the Black-Scholes model published by [13]. It is the most widely used model for estimating the price of European vanilla options. The model relies on the following assumptions:

1. The rate of return on the risk-less asset is constant;

2. The instantaneous log return of stock price is followed by GBM;

3. Stock does not pay a dividend;

4. There is no arbitrage opportunity in the market;

5. It is possible to borrow and lend any amount of cash at the risk-less rate, as well as buy and sell any amount of the stock;

6. No taxes and transactions cost included.

7. There is no limit to the scale.

The price of European call options, with strike K and maturity T can calculated by the Black–Scholes formula:

$$
\begin{aligned}
C(S,t) &= N(d_1)S_0 - N(d_2)Ke^{-r(T-t)} \\
d_1 &= \frac{ln(S_0/K) + (r+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \\
d_2 &= d_1 - \sigma\sqrt{T-t}
\end{aligned}
$$

The price of a corresponding put option based on put–call parity is:

$$P(S,t) = N(-d_2)Ke^{-r(T-t)} - N(-d_1)S$$

However, the Black–Scholes formula still has limitations. [19] found some evidence of inaccurate pricing by using the modified Black-Scholes model and option pricing model to study prices of European currency options traded in Geneva. As mentioned by [64], it appears that the rate of return on common stock can be treated as independent of the strike price and time to maturity, and follows a normal population with presumably constant mean but changing variance, a similar problem also realised by [56]. From the theory of volatility smiles and volatility term structure, [63] shows that the implied volatility is generally lower for at-the-money options but higher for out-of-the money and in-the-money options. In such cases, the stochastic feature of interest rates should not be ignored.

1.4.2 Time series model

Another class of the stochastic models in discrete time is the time series models. A time series is a sequence of data points over a time interval. Many sets of data are time series, such as daily stock price, monthly house price, yearly global temperature and the Cattle price index in Example 1. The impact of time series analysis on scientific applications, particularly in the field of economic and finance, as well as in physics and biology, has revolutionised the way people handle a sequence of data sets.

The primary objective of time series analysis is to develop mathematical models that provide proper descriptions of the sample data. Denote the value of a time series at t by y_t where $t \in (-\infty, \infty)$. Note t is typically discrete and varies over the integers. y_{t-1} indicates the value of y in the previous period and is called the first lag. One can introduce lag by considering the lag operator B where $y_{t-1} = B(y_t)$. Simply B^j

is the lag operator for iterated j times i.e., for its jth lag, we have $y_{t-j} = B^j(y_t)$, and the value of y that is h period ahead is denoted by $y_{t+h} = B^{-h}(y_t)$.

On the time-domain approach, as the landmark work of [16], develops a systematic class of models called ARMA models to explain the dependent structure of the current value on past values.

Definition 7. An ARMA process is defined as a combination of the auto-regressive process (of order p) and moving average process (of order q):

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

where ε_t is white noise process of random variables that are independent and identically distributed as:

$$\varepsilon_t \overset{iid}{\sim} \boldsymbol{N}(0, \sigma^2)$$

written in lag operator form gives:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Y_t = \delta + (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) \varepsilon_t$$

or

$$\phi(B)(Y_t - \mu) = \theta(B)\varepsilon_t$$

where μ is the unconditional mean that

$$\mu = \delta / (1 - \phi_1 - \phi_2 - \cdots - \phi_p)$$

and $\phi(B)$ and $\theta(B)$ are called lag polynomials such that

$$\begin{aligned} \phi(B) &= (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) \\ \theta(B) &= (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) \end{aligned}$$

Definition 8. The ARMA process is stationary if the roots of

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

are outside the unit circle; then, we can obtain:

$$Y_t = \mu + \psi(B)\varepsilon_t$$

where

$$\psi(B) = \frac{(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)}{(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)}, \sum_{j=0}^{\infty} |\psi_j| < \infty$$

and ψ_j is the polynomial for the infinite-order moving average representation of the ARMA process, see [43] for details.

To simplify the model, in the following we assume that the mean value $\mu = 0$. In cases where the AR component of the process Y_t has a unit root, i.e. $z = 1$, and all other characteristic roots are outside the unit circle, we can write the non-stationary process Y_t as

$$\phi(B)Y_t = \phi^*(B)(1 - B)Y_t = \theta(B)\varepsilon_t$$

then

$$\phi^*(B)\Delta Y_t = \theta(B)\varepsilon_t$$

Since all the roots of $\phi^*(z) = 0$ are outside the unit circle, the polynomial of $\phi^*(B)$ is invertible; then, the first difference of the process $\Delta Y_t = Y_t - Y_{t-1}$ is stationary. In such a case, the process for Y_t is said to be first difference stationary, or difference stationary.

If a process Y_t has one and only one unit root, and becomes stationary after first differencing, it is said to be integrated of order one, denoted by $I(1)$. Moreover,

Definition 9. If process Y_t's first difference ΔY_t is described by a stationary ARMA(p,q) process, then process of Y_t called the Auto-regressive Integrated Moving Average (ARIMA) process of order $p, 1, q$, denoted by ARIMA (p,1,q).

In general, for an ARMA(p,q) model:

$$\phi(B)Y_t = \theta(B)\varepsilon_t$$

If the process has d unit roots, and become stationary after taking dth difference, then Y_t is said to be integrated of order d.

$$\phi^*(B)\Delta^d Y_t = \theta(B)\varepsilon_t$$

Thus, a non-stationary ARMA(p,q) model with d unit roots is equivalent to an ARIMA(p-d,d,q) model for Y_t, and can be rewritten as:

$$\Delta^d Y_t = \frac{\theta(B)}{\phi^*(B)}\epsilon_t$$

The ARIMA process has been widely used for modelling and prediction purposes. For instance, [72] used the ARIMA model for baseline building blocks for impact assessment, forecasting, and causal modelling. [43] proposed this type of model for demonstrating changes in regimes based on U.S. real GNP data, which can used as an objective criterion for measuring U.S. economic cycles.

We shall introduce the Autoregressive Fractionally Integrated Moving Average (ARFIMA) from [40] in Chapter 3, compare to ARIMA modelsm, ARFIMA model can better modelling long-memory time series processes.

1.5 Problems descriptions

1.5.1 Frost losses and insurances

Chapter 2 introduces a possible solution to Question 3 by introducing a quantitative risk management method to manage the risk of frost losses (see [8]). Agricultural insurance is a type of risk managerial tool used by farmers to hedge against agricultural losses. The main risks in the agricultural industry are price risk and production risk. While the former is caused by price volatility due to mismatches between demand and supply, the latter is usually caused by adverse natural events. According to [51], agricultural insurances can be classified into three major groups: Indemnity-Based Agricultural Insurances, where the losses are the main factors for the claims; Index-based Agricultural Insurances, where the claims are based on some indexes; and Crop Revenue/Yield or Crop Insurances, that guarantee a certain level of return to the commodity prices. For further reading on agricultural insurances, please see [65], [115], [116], [77], [48], and [14].

Even though the well-established concept of frost insurance has extensively been used in practice, there are only a few papers in the literature focusing on this. For instance, [59] developed a temperature insurance scheme as an alternative to direct frost damage insurance, [90] addressed the damage from frost and other weather disruptors, focusing on risk transfers rather than studying the frost insurances, and [89] studied the risk management instruments for frost irrigation and hail nets. On the other hand, since frost is an event caused by low temperature, the other most relevant literature on frost insurances concerns weather index-based insurances. Starting by [106], the author argues that, even though the weather derivatives can be used as a form of agricultural insurances, location specifications should be adopted in their pricing. [30] show the efficiency of temperature-humidity index insurance, whereas in [29] the authors evaluated three index-based crop insurance contracts in the same area. In [108], a web-based computer program is designed to evaluate weather risk. The demand for weather-based insurance policies for rural areas is studied in [76]. While [58] compared the efficiency of various weather indexes, [103] compared different methods for pricing weather derivatives based on growing degree-days. In addition, studies on systemic risk in index-based crop insurances in an equilibrium pricing framework can be found in [17] and [94].

Nonetheless, we see an extremely small amount of information on the report for frost damages; the only reliable data source we have is the Florida citrus freeze fact sheet from Florida Citrus Mutual. For an event that happens once over the last ten years and three times in the last twenty years in this area, the traditional risk-oriented pricing method does not work well due to a lack of observable data points. Thus, a creation of a quantitative method is needed here to find the fair price of protection.

The main challenges in modelling a frost-loss variable based on temperature are, first, the losses are a non-linear function of temperature, and, second, crops can resist the first few hours of freezing temperature. In Chapter 2, a universal frost damage model based on temperature, regardless of any frost protective method, is proposed. The model bases its fundamentals on a loss function introduced by [109] and a temperature loss magnitude model by [99].

After characterising a model for our loss variable by finding its cumulative distribution function, we use distortion risk premiums to price a contract that does not raise the risk of moral hazard. Distortion risk premiums are readily used for pricing insurance and reinsurance contracts in the literature of actuarial science; see [21], [78] and [3] for reference. In particular, [78] used some distortion risk premiums for designing and pricing optimal reinsurance on agricultural goods.

Meanwhile, the risk of moral hazard is one of the most important issues discussed in pricing agricultural insurances. It occurs when the risk of losses is not felt by one of the parties, either the farmer or the insurance company. We benefit from a method in the corresponding literature that removes the risk of moral hazard by considering that contracts are non-decreasing functions of the total risk. A particular example of such contracts is a two-layer stop-loss policy that is introduced and used in Chapter 2.

The prices of two-layer stop-loss policies on frost losses for the citrus fruits will be studied, using Value at Risk, Conditional Value at Risk and Wang's premium based on a data set from San Joaquin Drainage, California (Figure 1.5.1).

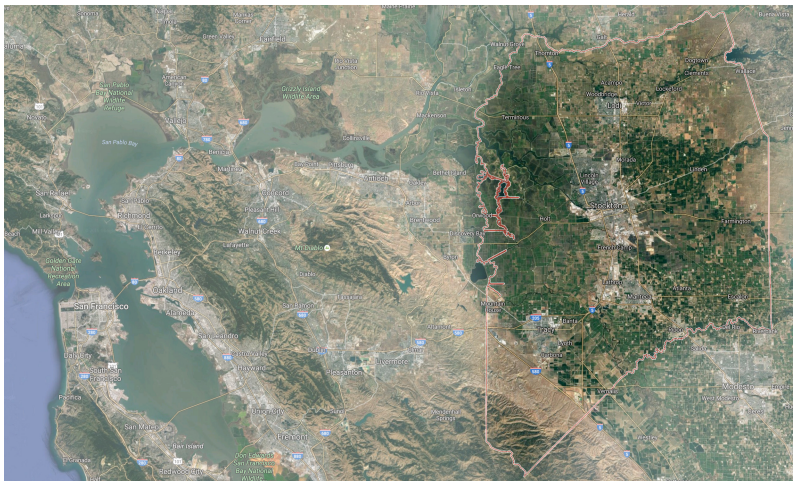


Figure 1.5.1: Satellite photo for San Joaquin(Wah 'Keen) in CA, US (source: Google Earth)

1.5.2 Modelling long memories[1]

Question 4 can be partially answered by using traditional time series model including the short memory ARIMA processes. However, in our study of real commodity data, we find the Box-Jenkins method on the selection of ARIMA model does not work very well. This is partially due to the limited capacity for the ARIMA model to capture the long memory property of the underlying time series. Figure 1.5.2 shows the auto-correlation function and partial auto-correlation function for the Cattle index (from example 1, the first difference is taken to reach stationarity) sample to the lag of 50. One can see, both correlations failed to converge to 5 critical levels indicated by the blue lines. Figure 3.3.7 shows such a problem exists in most of the products we studied.

[1]This section focuses on the problem introduced in Chapter 3

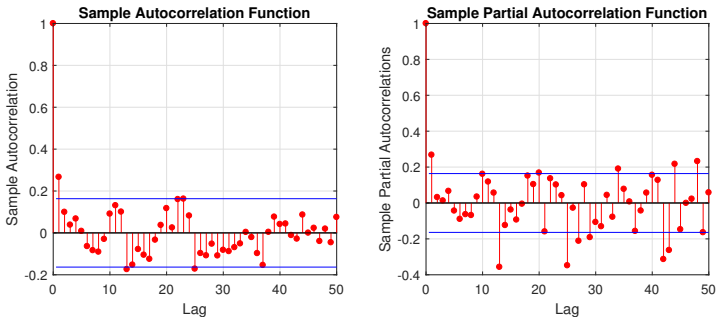


Figure 1.5.2: Sample auto-correlation and partial auto-correlation plots for Cattle index data after first difference

Indeed, LRD or long memory is an important phenomenon in finance and economic time series, see [61] and [88]. Until suitable stochastic processes became available in the literature, deviations from the independence assumption were considered far beyond usual modelling practices. However, ground-breaking work by Mandelbrot (and his colleagues) in introducing "fractals" and 'self-similarity' changed the paradigm, [66, 70, 69], essentially developing a new branch of mathematics for studying LRD, particularly based on scaling and fractal behaviour (see [67]).

Long-range dependent processes are characterised by a time series' auto-covariance function. In a LRD processes, the decay of the auto-covariance function follows a power law, decaying slower than the normal exponential function which has a shorter memory. Most of the long-range dependence definitions can be found in the literature based on the second-order properties of a stochastic process. This is mainly due to the fact that the second-order properties are conceptually simpler and easier to estimate. A standard definition of linear dependence structures is given as in [12].

An elegant way to study LRD is to estimate the Hurst exponent which measures the persistence or mean reversion of a time series. There are many obstacles to using the Hurst exponent in studying LRD processes. For instance, in estimating the Hurst exponent, there are a few assumptions that one needs to make on the distribution of the underlying time series without which the existing estimation methods of the Hurst exponent are usually unstable. In order to overcome this problem, in Chapter 3, by adopting an iterative convergence approach, we develop a new Hurst coefficient estimator, which is more reliable than the existing methods in the literature and test them on simulated data. Then, we use this method to estimate the Hurst exponent

of the commodity data where we observe that under certain measures, Hurst values distinguish agriculture and non-agricultural products.

1.5.3 Feasibility analysis for a new type of agriculture insurance

In Chapter 4, we will analyse the feasibility of a new type of price-based agriculture insurance in the UK agricultural markets.

According to the United States Department of Agriculture[2], five factors are described as general risks for the farming businesses.

Production risks from weather, diseases, pests and other production related factors are fundamental risks for the farmers. As we will discuss in Chapter 2, this type of risk can be avoided by physical protections applied to the production process, or compensated by the purchasing of Crop insurance, a type of risk management tool to protect farmers against production risks for agricultural commodities, see [44, 52, 93] and two papers from Miranda [73, 74] for details.

Marketing Risk and Financial Risk are listed as the second- and third-most important risks for agriculture goods. These risks represent the scenarios where farmers and producers may suffer from the lower- or higher-than-expected product prices, or have insufficient profits to allow the production to continue. Part of the reason for these risks is the commodities prices' volatilities. According to [37], the nature of commodity is more volatile and unstable when compared to other financial or physical products. This will be discussed in Chapter 4 with comparisons using Sharpe Ratios. Long debates have happened in the industry for the causes of this phenomenon, and so does the justification for the existence of commodity trading market. Even though the public trading commodity market only accounts for one-fifth of the total global trading volumes, some people still blame the market players' speculations as the cause of the high volatilities in commodity prices.

Nevertheless, in the mainstream literature, the general attitude towards to the existence of such markets like CBOT, Euronext and the multiple marker players like arbitragers, brokers and speculators is that these help to reduce the uncertainty of

[2]USDA, https://www.ers.usda.gov

commodity prices, providing the necessary liquidity to the market itself[3]. Also, different expectations for future commodities prices will help large agricultural producers to reduce and hedge against their risk exposures. Unfortunately, such types of risk protections are not available for most of ordinary farmers. Entrance barriers such as large deposit requirements, technical trading skills and perhaps regulations together make the accessing to the financial commodity market almost impossible. As a result, farmers can only choose a handful of risk management tools to help them reduce these risks. Some of the traditional methods to prevent Marketing and Financial risks involve actions such as joining marketing cooperatives, diversifying productions, delaying harvests or even developing strategic business plans. The general lack of an easy and accessible financial tool for the ordinary farmer to hedge against these risks inspired us to create a new type of insurance. In Chapter 4 we will provide a possible solution to these problems by providing a new type of insurance based on the price of commodities only. Thus, marketing risk and financial risks can be hedged directly by purchasing prices protection, avoided by the compensations from underwriters in not-desired scenarios whether it is a price fall for producers or price raise for wholesalers. When compared to traditional crop yield or revenue insurance, our policies only rely on third-party or a trusted public price index data, so problems like measuring the yield, preventing moral hazard and lacking empirical data are avoided (see [51] and also [73] for more details about these risks).

[3]We do not necessarily advocate this idea.

1.6 Our contributions

In this thesis, we study some aspects of risk management in insurance by answering the five questions posed in Chapter 1 in the next four Chapters.

In Chapter 2, we propose to price frost insurances with financial quantitative methods; this is to answer Question 3. The aim is to design and price insurances on frost damages to protect the farmer. The main challenge here is that the data on frost are so insufficient that it is almost impossible to propose a proper model. The model we proposed manages the risks from the formation and original, by using the temperature data set, which is widely accessible, and we hope to reduce risks from the places where it starts.

Compared to the existing literature, we see our work as unique in two ways. First, the temperature data are used to construct a model for losses caused by frost, and, from that perspective, we choose not to use temperature just as an index (while it is an index). Indeed, from an insurance pricing perspective, the contracts that we price can be considered as adjusted index-based insurances. However, here the damage is based on temperature that by using a non-linear loss variable, can properly replicate the real losses. To further explain our point, note that an important issue with an index-based insurance is basis risk. Basis risk in index-based insurances arises when the index measurements do not match an individual's actual losses. The basis risk can be a result of a poorly designed product or geographical effects. Our approach can help mitigate the first issue. The basis risk could be tested by comparing the payoff of the proposed insurance and the actual production losses; however, that is out of scope of this thesis, and is left as an important and interesting area for future research.

Second, our loss variable model is obtained by using the temperature data inspired by the engineering method from [59]. There are several reasons why we do not use the data of frost events to directly construct a loss variable. The first reason is that the data of frost events are usually non-existent, hardly available or not suitable to construct a model, making us incapable of carrying out reliable statistical tests. This problem can also result in overseeing major losses and, in particular, the rare and catastrophic events. The second reason is that the frost damages data cannot exactly reflect the losses caused by frost since the farmers usually try to manage the risk

of frost by getting involved in some physical protective activities such as irrigation, pruning, heating, etc. That is why the damages caused by frost events can be strongly dependent on the method used in a particular area.

In Chapter 4, Question 4 is discussed on modelling process with long memory. Our contributions to the existing literature are, first, we proposed a stable parameter estimation method for $ARFIMA$ processes; second, the method is applied on a set of commodity data wherein their best ARFIMA fittings are found; third, we discovered that, for most of the commodities, there exists long-memory, and a distinct trend is observed from products with different demand elastics.

In the existing literature, there are mainly three ways to estimate the parameters on an ARFIMA model, maximum likelihood methods, semi-parametric methods and the heuristic approach. While the first two method all rely on the likelihood functions which are difficult to find due to parameter mis-identification, the third approach, heuristic, simplifies the process by hiring a mathematical relation between Hurst exponent and the fractional parameter in the ARFIMA process. The recursive method that we propose in Chapter 3 stabilises the estimation method for the Hurts exponent and yields the best estimations when compared to other methods.

Also, real market data are employed in this research. By studying the absolute of the log return and the increments data over 40 trading contracts over the last 20-30 years, we discovered that, for all the increments, data (first difference) have a long memory in present, and most of the absolute log returns also exhibit long memory. This discovery is consistent with the experience from the industry as well as previous studies in Hurst and the elasticity of product demands.

Another new observation in Chapter 3 is the Hurst values and their corresponding products when incremental data are of interest. We observe an increasing trend from Live animals to crops and then to Cattle data.

The contribution from Chapter 3 is to provide a method to find the parameters for a powerful long memory model. This model can help us to model the past and predict the future better; these eventually can help us to better simulate and control the existence of risks.

Chapter 4 tries to review Question 3 and answer 5. In the commodity risk management, the demand side usually consists of producers and wholesalers, such as,

farmers, food factories and supermarkets. They need risk management tools to protect against adverse events like price fluctuation, loss on harvest due to bad weather or pests, policy changes like Brexit etc. Insurance policies can be used in almost all of those events to hedge against potential losses.

Different from Chapter 2 where the direct loss variable is measured based on harvest - in that case, damages from frost - we will present agriculture risk management from a new perspective. By considering the demand side in agricultural risk management, we come up with an optimal solution that minimises insurers' global risk positions, i.e. both the risk exposures and the premiums. Also, for the supply side (these are usually underwriters, risk capitals), an investment strategy is constructed with two-layer stop-loss style insurances. A portfolio of such strategies is then bench-marked with main financial indices to show their feasibility and profitability of them.

More specifically, demand side faces risks like 1. exposure that are not covered by the insurance contract[4] and 2. high premiums paid to purchase insurances. If proper risk measures are applied to both risks, one can try to minimise the global risk by altering different policy types, e.g. quote-share or stop-loss policy. We shall provide a theoretical solution to this problem in the first half of Chapter 4 together with sensitivity analyses under several market assumptions. As a result, we find that a two-layer insurance policy exists to solve this problem if certain conditions are met.

On the other hand, the underwriters and risk capitals are usually the supply side of these risk protection tools. From a supply point of view, a two-layer stop-loss style insurance also makes sense in terms of risk management requirements. In the second half of Chapter 4, we shall discuss the spread-style hedging strategies for the underwriters. These contracts enable the supply side to provide risk protection to the demand side, as well as protect themselves from adverse events that can cause extreme losses. After a short introduction of the contracts, we will discuss a modified pricing method based on [5], [4]. Such a pricing method considers suppliers' risk appetites and generates a corresponding premium that fits suppliers' requirements. We found that, based on 10 UK agriculture product data sets, such two layer polices

[4] (for example, the loss higher than the lower retention levels in a stop-loss insurance)

are feasible and even good investments for the suppliers based on the Sharpe Ratio analysis.

Chapter 2

Frost Insurance and Reinsurance

The main objective of this chapter is to model the losses caused by frost events and use it to price frost insurances. Since the data on frost events are either unavailable or rarely available, we have chosen to obtain a model for frost losses based on temperature by using some fundamental agricultural engineering findings on frost damages. The main challenges in modelling frost loss variables are first, the non-linearity of the frost losses with respect to the temperature and second, the fruit resistance to the first few hours of low temperature. We address both issues when introducing our frost loss variable. Then after finding the loss model, we use it to price frost insurances for a general family of insurance contracts that do not generate any risk of moral hazard. In particular, we will find the premiums of stop-loss policies for losses to citrus fruits using Value at Risk, Conditional Value at Risk and Wang's premium based on temperature data from San Joaquin Drainage County in California.

2.1 Moral hazard

In an insurance contract, moral hazard happens in a scenario where one party does not feel the risk of losses. [1] is among the first who realised that adverse selection and moral hazard are important phenomenons in finance and insurance. According to [28], the concept of moral hazard has been used for more than 200 years and originates from insurance and economic decision-making literature. [102] uses a game theoretical approach to show how the competitive equilibrium contract is used as a solution to sharecropping in presence of moral hazard. Later [101] does a comprehensive research on the connections between risk, insurance, incentives and imperfect information. In

agriculture, [46] examine the U.S. mid west data for crop insurance using some utility functions in presence of moral hazard risk in the farmers businesses. They suggest fertiliser and pesticides may be risk-increasing inputs as they find farmers with insurance protections use more of them. [98] look at the Kansas dry-land wheat farmers using econometric methods and come up with a different conclusions. They find that insured farmers use fewer chemicals in the field due to moral hazard incentives. [114] again determines the crop insurances lead to an increasing use of chemicals due to moral hazard risk. He also studies the crop mix problem when designing crop insurance. Finally, [85] employ a new linear model to estimate the extent for which the moral hazard in crop insurances exists. Interestingly, [112] use a new approach on simulations with moral hazard costs and conclude that, moral hazard costs can even be removed or reduced by the design of insurance itself.

In the present thesis, we chose an approach where the risk of moral hazard is reduced by setting contracts that both parties, the insurer and the insuree, feel the losses. We assume that the market's attitude is toward avoiding the moral hazard risk. Therefore, the contracts have to move co-monotonically with the total risk. This approach is widely used in the literature of actuarial mathematics e.g., [36], [18], [21], [20], [22], [3] and [117].

If a contract X is a function f of the loss variable Y, i.e., $X = f(Y)$, by assuming that both the insurance company and the farmer have to bear the risk generated by the loss variable Y, both insurance and farmer's loss variables, $f(Y)$ and $Y - f(Y)$, have to be non-decreasing functions of Y. Therefore, we have to assume that the functions $x \mapsto f(x)$ and $x \mapsto x - f(x)$ are non-negative and non-decreasing real functions. To set this economic assumption on a sound mathematical basis, the following set of admissible contracts is introduced by

$$\mathcal{A} = \{f(Y) | f \in C\},$$

where

$$C = \{f \geq 0 | f(x) \text{ and } x - f(x) \text{ are non-decreasing}\}.$$

Note that all members in C are Lipschitz continuous since for any $0 \leq x_1 \leq x_2$ we have $x_1 - f(x_1) \leq x_2 - f(x_2)$ and $f(x_1) \leq f(x_2)$, which result in $0 \leq f(x_2) - f(x_1) \leq x_2 - x_1$. In particular, we have that $0 \leq f(x) \leq x$, $\forall f \in C,\ x \geq 0$. It is known

that every Lipschitz continuous function f is almost everywhere differentiable and its derivative is essentially bounded by its Lipschitz constant (see Theorem 7.20 in [91]). Furthermore, f can be written as the integral of its derivative denoted by h, i.e., $f(x) = \int_0^x h(t)dt$.

Definition 10. For any indemnification function $f \in C$, the associated marginal indemnification is a function $0 \leq h \leq 1$ such that

$$f(x) = \int_0^x h(s)ds,\ x \geq 0.$$

The following proposition is very helpful to compute the premium of a contract when we have the marginal indemnity.

Lemma 1. *Let us assume that π is a distortion risk measure. If $X = f(Y)$ where $f(x) = \int_0^x h(t)dt \in C$, then*

$$\pi(f(X)) = \int_0^\infty g\left(\bar{F}_X(t)\right) h(t)\, dt,$$

where $\bar{F}_X$ is the survival function of X.

Proof. See Lemma 2.1 in [117]. □

In this chapter, particularly we are interested in two-layer stop-loss policies which are popular in agricultural insurance industry.

Definition 11. A two-layer stop-loss contract, f_{stop}, with a deduction level r and a retention level $c > r \geq 0$ is as follows

$$f_{stop}(x) = \begin{cases} 0, & x \leq r \\ x - r, & r < x < c \\ c - r & x \geq c \end{cases}. \tag{2.1.1}$$

It follows that in this case $h(x) = 1_{[r,c]}(x)$, which results in the following premium

$$\pi(f(X)) = \int_0^\infty g\left(\bar{F}_X(t)\right) h(t)\, dt. \tag{2.1.2}$$

$$= \int_r^c g\left(\bar{F}_X(t)\right) dt. \tag{2.1.3}$$

2.2 Frost insurance

2.2.1 The proposed model

In the volume two of the Food and Agriculture Organization (FAO) book on frost protection, [42]'s loss model for the frost risk management is used, where the loss variable is modeled simply by a Bernoulli distribution. In mathematical terms, [42] denotes the probability of having frost event over a year by p_f, and finds the probability of having at least one frost event over n years by $R = 1 - (1 - p_f)^n$. He also suggests that the probability can be found by considering a type I extreme distribution for the frost temperature, i.e.,

$$p_f = 1 - \exp\left(-\exp\left(\frac{T_c - m + 0.45s}{0.78s}\right)\right), \tag{2.2.1}$$

where T_c denotes the frost temperature, m is the average temperature and s is the empirical standard deviation for temperature over a one-year duration.

This model is not rich enough for our purposes since first, it does not specify the amount of damages and second, it does not specify how long after freezing temperature the crop starts to get damaged. That is why we have to study further the fundamentals of the frost damage modelling and temperature forecasting, and construct a more sophisticated model.

Consequently, we are inspired by the approach introduced by [109] for measuring the risk of frost damages. First of all, we need to introduce some necessary notations. Let us denote the total production without any loss by A. Then, one can introduce the total loss by $Y = A \cdot \mathbb{D}$, where $\mathbb{D}$ is the percentage of the production loss. In the following, we will give a model for $\mathbb{D}$. As before, let us denote the frost temperature at which the damages will start by T_c, and the temperature at which the total crop is damaged by T_t. According to [109] the percentage damage at temperature T, denote as $D_d(T)$, for a long enough freezing time interval, is defined as follows

$$D_d(T) := \begin{cases} 0 & \text{if } T > T_c \\ \left(\frac{T - T_c}{T_t - T_c}\right)^{\eta} & \text{if } T_t \leq T \leq T_c \\ 1 & \text{if } T < T_t \end{cases}, \tag{2.2.2}$$

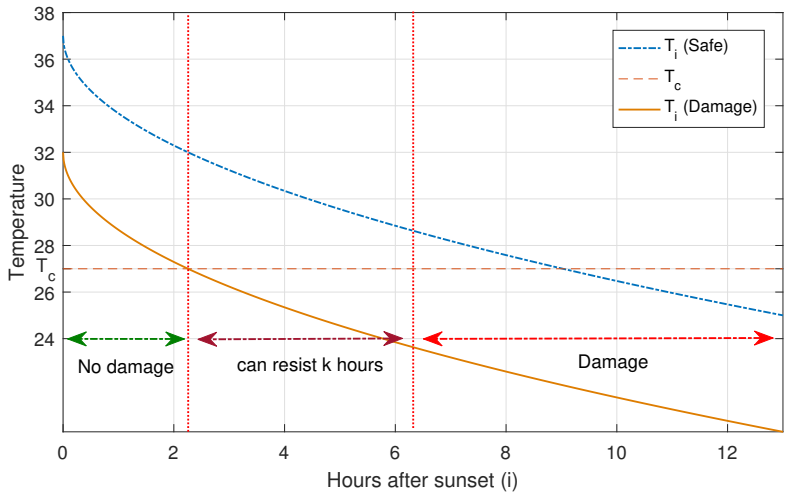


Figure 2.2.1: Temperature after sunset before sunrise, illustration of the damage mechanism with respect to temperature and time.

where $\eta \geq 1$ measures the vulnerability rate of the crop to the low temperature. For instance, it is reported that for citrus fruits $T_c = 27°F$[1], $T_t = 20°F$ and $\eta = 1.5$, however, the crop can resist few hours of freezing temperature before being damaged, for instance, orange takes at least four hours below $T_c = 27°$ to have a damage ([109]). Let us assume that the crop is damaged only if the temperature is k hours below T_c, i.e., the adverse event occurs if $T < T_c$ last at least for k hours. To know the number of hours below T_c, we use a square root model by [99] which can forecast the temperature of the weather over a single night. If we denote the temperature at sunset by T_0, the temperature i hours after the sunset and before the sunrise by T_i, and finally the number of hours between sunset and sunrise by I, the temperature i hours after sunset is given by

$$T_i = T_0 + (T_I - T_0)\sqrt{\frac{i}{I}}\ ,\ i \in [0, I]. \tag{2.2.3}$$

Figure 2.2.1 demonstrates the model using two different scenarios; one with higher and one with lower sunset temperature. Given that the temperature at sunrise is always lower than sunset (i.e., $T_I < T_0$), then the temperature is a decreasing function of time. The temperature decreases at a square-root speed, it is convex and tends to change more slowly towards sunrise. This illustration generated by model is consistent

[1]Note that °F = Degree Fahrenheit.

with the data observations which we use in the next section for calibrations. Since T_i is decreasing in i, this simply suggests that damages occur if $T_{I-k} \leq T_c$. Based on this and the definition of the damaged function $D(T)$, we can introduce the percentage of damages as follows

$$\mathbb{D} = D_d(T_I) 1_{\{T_{I-k} \leq T_c\}}. \tag{2.2.4}$$

Next, we find the distribution of the loss variable $\mathbb{D}$. Let $d \in [0,1]$, we have to find $\mathbb{P}(\mathbb{D} \leq d)$. But before that we need to introduce some notations and consider some assumptions. Let $b = T_c - T_t$, $B = \sqrt{\frac{I-k}{I}}$, $M = T_0 - T_c$ and $N = T_0 - T_I$. In the sequel, we assume that T_0 and $T_0 - T_I$ are independent, which results in the independence of N and M. This assumption will be justified with the real data in the next section. Note that since, $T_0 \geq T_I$, we always have $N \geq 0$. Furthermore, we assume that $F_N(0) = 0$ and that N and M have continuous distributions. Additionally, if $d = 1$ clearly we have $\mathbb{P}(\mathbb{D} \leq 1) = 1$. So let us assume $d \in [0,1)$. We have the following proposition

Proposition 1. *Let us introduce the following function*

$$\mathcal{H}(x) := \int_{-\infty}^{\frac{Bbx}{1-B}} \bar{F}_N(bx+z) f_M(z)\, dz + \int_{\frac{Bbx}{1-B}}^{+\infty} \bar{F}_N\left(\frac{z}{B}\right) f_M(z)\, dz,$$

if we assume M and N are independent then we have

$$\mathbb{P}(\mathbb{D} > d) = \mathcal{H}\left(d^{\frac{1}{\eta}}\right) - \mathcal{H}(1), \tag{2.2.5}$$

where $\bar{F}$ is the survival function and f is the PDF.

Proof. See the Appendix A.1. □

2.2.2 Two analytic solutions

There are not a lot of scenarios that one can find an analytic form for the distribution of $\mathbb{D}$. Here we present two of them when N is exponential while M is either normal or exponential distributed.

Proposition 2. *Assume $N \sim exp(1/\mu_N)$ and $M \sim \mathcal{N}(\mu_M, \sigma_M)$ are independent. Then*

$$F_{\mathbb{D}}(d) = 1 + \mathcal{H}(1) - \mathcal{H}\left(d^{\frac{1}{\eta}}\right), \tag{2.2.6}$$

where

$$\mathcal{H}(x) = \exp\left(\frac{\sigma_M^2}{2\mu_N^2} - \frac{(bx+\mu_M)}{\mu_N}\right)\Phi\left(\frac{\frac{Bbx}{1-B} - \left(\mu_M - \frac{\sigma_M^2}{\mu_N}\right)}{\sigma_M}\right) + \exp\left(\frac{\sigma_M^2}{2B^2\mu_N^2} - \frac{\mu_M}{B\mu_N}\right)\Phi\left(\frac{\mu_M - \frac{\sigma_M^2}{B\mu_N} - \frac{Bbx}{1-B}}{\sigma_M}\right).$$

and Φ *is the CDF for the standard normal distribution.*

Proof. See the Appendix A.2. □

In the following we find the CDF of $\mathbb{D}$ when M follows an exponential distribution.

Proposition 3. *Assume* $N \sim exp(1/\mu_N)$ *and* $M \sim exp(1/\mu_M)$ *are independent. Then*

$$F_{\mathbb{D}}(d) = 1 + \mathcal{H}(1) - \mathcal{H}\left(d^{\frac{1}{\eta}}\right), \tag{2.2.7}$$

where

$$\mathcal{H}(x) = \frac{\mu_N \exp\left(-\frac{bx}{\mu_N}\right)}{\mu_N + \mu_M}\left(1 - \exp\left(-\left(\frac{1}{\mu_N} + \frac{1}{\mu_M}\right)\frac{Bbx}{1-B}\right)\right) + \frac{B\mu_N}{B\mu_N + \mu_M}\exp\left(-\left(\frac{1}{B\mu_N} + \frac{1}{\mu_M}\right)\frac{Bbx}{1-B}\right)$$

Proof. See the Appendix A.3. □

Note that

$$D_t(T_I) = \left(\frac{T_I - T_c}{T_t - T_c}\right)^{\eta} 1_{\{T_t \le T_I \le T_c\}} = \left(\frac{N-M}{b}\right)^{\eta} 1_{\left\{1 \ge \frac{N-M}{b} \ge 0\right\}}. \tag{2.2.8}$$

From Eq. (2.2.8), one can see that if M takes larger negative values, for example if M is normally distributed, then the risk will decrease. In contrast, if M is exponentially distributed, then we have a riskier situation. Therefore, we always expect the insurance prices for an exponential M to be larger than a normal M.

2.3 Calibration for citrus frost insurance contracts

In the following, we price a two-layer stop-loss crop insurance for citrus fruits. A study in the Florida citrus freeze fact sheet[2] shows that weather conditions prior to cold temperature, duration of cold, position of the trees in the grove or yard, maturity of the fruits, health and age of the trees can affect trees and fruits hardiness. Usually, when the temperature falls below $28^\circ F$ for more than four hours, fruits start to be damaged. Four hours at $20^\circ F$ can kill $3/8$ inch or smaller wood and temperatures below $28^\circ F$ for 12 continuous hours may kill larger limbs and possibly the entire tree. On the other hand, [109] mention that based on an experiment conducted by the University of Florida, the value for η for orange is reported to be equal to 1.5. The same experiment confirms that an orange needs to stay at least four hours below $28^\circ F$ to be damaged. Consequentially, we consider the following set of parameters for our study

- $k = 4$,
- $T_c = 28^\circ F$, $T_t = 20^\circ F$, which results in $b = T_c - T_t = 28 - 20 = 8$ $(^\circ F)$;
- $\eta = 1.5$.
- $I = 14, 13.75$ and 13 separately for December, January and February based on monthly average

According to [99], there are two main protection methods: active and passive. While passive protection methods are done in advance of a frost night; for example, careful site selection, managing cold air drainage, careful plant selection, proper pruning, using plant covers etc., active protection methods are done during a frost night; for example, using heaters and wind machines, helicopters, sprinklers, surface irrigation etc. A proper insurance contract, which should not increase the risk of moral hazard, encourages all protection methods. Here, we consider a stop-loss insurance contract which covers $c \times 100$ percent of the citrus losses caused by frost and low temperature. In this particular contract, since a farmer needs to cover all losses above $c \times 100$ percent, he is encouraged to passively and actively be involved in the crop protection.

[2]Provided on-line by Florida Citrus Mutual.

	December		January		February	
	A_p	K_p	A_p	K_p	A_p	K_p
M	0.0010	0.1000	0.0010	0.0100	0.0010	0.0916
N	0.0010	0.1000	0.0010	0.1000	0.0010	0.1000

Table 2.3.1: p-values for ADF and KPSS tests for stationary.

We are interested in pricing monthly contracts, thus monthly temperature data of San Joaquin Drainage County in the state of California, US, is used, where there are large groves of citrus located[3]. The data is taken from National Oceanic and Atmospheric Administration (NOAA) online data set arranging from 1895 to 2016, a total of 122 years. The damages to the citrus production usually happen during December, January, and February, when there is a risk of frost. Here, consider the data of the minimum and the average weather temperature in our studies: T_I is the minimum temperature and T_0 is the average temperature. Note that T_0 is assumed to be the temperature at sunset, however, since this data is not available, we use the average temperature data of each day instead. Furthermore, since we require the stationarity of the data, we cannot consider the time series data over the whole year, rather, we fix into three months (hence January, February or December), and study the time series data of each month over the last 122 years.

Before the independence and goodness-of-fit tests for M and N to construct suitable models, the stationarity tests of the time series are carrying out by the augmented Dickey-Fuller (ADF) and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests[4]. The p-values of the tests are denoted by A_p and K_p, respectively. While the former has the null hypothesis of having a unit root, the latter uses stationarity around a deterministic trend as null hypothesis. One can see in Table 2.3.1 that our ADF test rejects all the null hypothesis with very small p-values. The KPSS test fails to reject the stationarity except for January's M data, but with a relative small p-values. Thus, we assume all our data is stationary and suitable for fitting purposes.

Next, the serial correlations and stationarity for each individual M and N are tested. Figure 2.3.1 presents the auto-correlations for all 6 sets of data from December to February. One can tell, nearly all correlations at time lags 2 to 20 are below the 0.05

[3]Our frost damage model could be applied to any other places, as our model is based on temperature data set and this data set is accessible for almost every other place.

[4]Note, for ADF test, we have removed the deterministic trend from the data. The minimum and maximum p-values are set at 0.001 and 0.1.

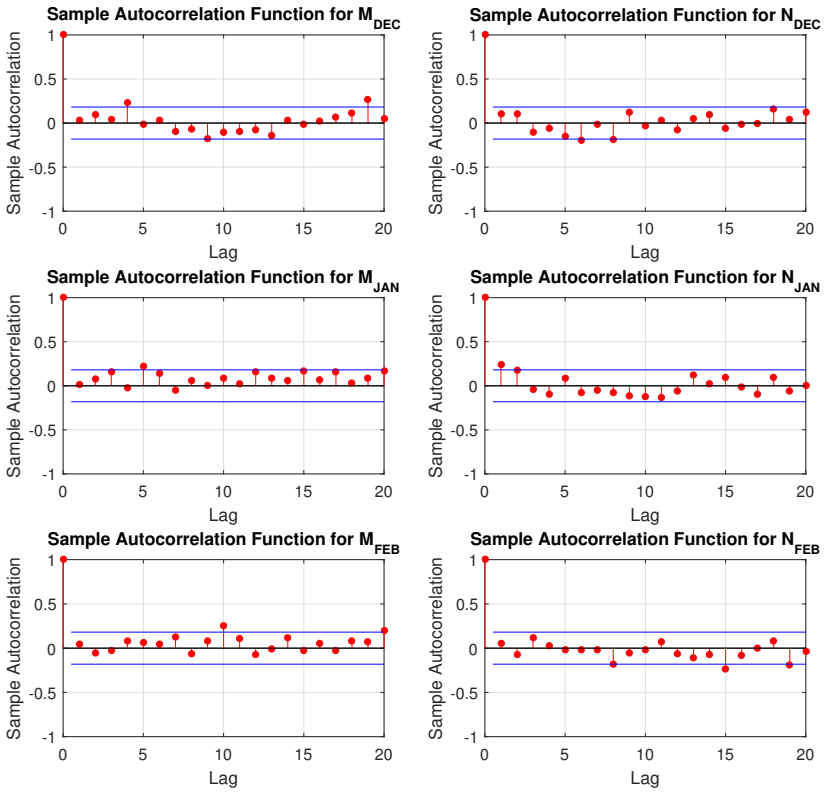


Figure 2.3.1: Auto-correlations for M and N in December, January and February.

critical values which are marked by the blue lines. Since we have 122 data points for each month, we believe testing up to time lag of 20 is adequate. The auto-correlation plots rule out the necessity of employing time series models in this particular study. Thus, in what follows, we focus on finding reasonable CDFs for M and N.

To justify the independence assumption, in Table 2.3.2, the Pearson product-moment correlation coefficient and the distance correlation coefficient, as suggested by [104], are obtained for M and N from December to February. Together, their p-values are also presented to show the level of independence as it is assumed. One can tell, the low correlation coefficients indicate lack of correlation between M and N for all three months, particularly from the distance correlation coefficient, since it only equals to 0 if and only if M and N are independent. The relatively large p-values shows that at the 5% significance level, the evidence from data is not against the null hypothesis of independence, so our assumption that M and N are independent is reasonable.

		Test Statistics	p-values
Dec	Correlation	0.1189	0.1941
	Distance Correlation	0.1676	0.2687
Jan	Correlation	0.0299	0.7437
	Distance Correlation	0.1660	0.3819
Feb	Correlation	0.1021	0.2630
	Distance Correlation	0.1499	0.4939

Table 2.3.2: Independent test for M and N.

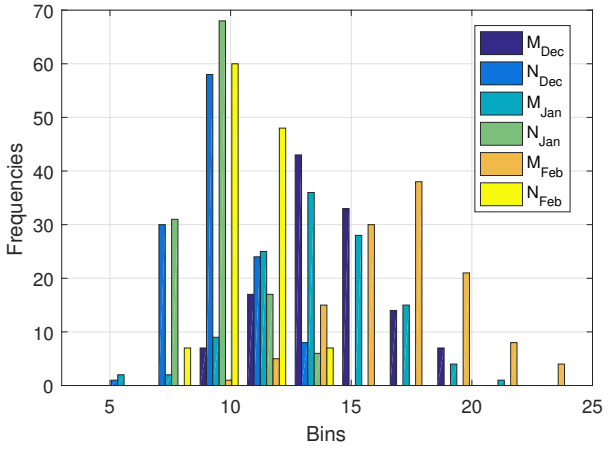


Figure 2.3.2: Histogram for all M's and N's with bin size of 10.

A distribution list from the exponential family is chosen after observing the histograms for all M's and N's in Figure 2.3.2. In the next step, we mange to fit all 6 sets of data into Normal, Gamma, Weibull, Log-Normal and Exponential densities and test the goodness-of-fit by carrying out a simulation study.

The approach we take for finding relatively realistic distribution functions of all M's and N's can be explained as follows. We first fit M and N into the distribution list as in Table 2.3.3 using method of moments[5], then, we generate 500 sets of simulated data with the fitting results for each M and N. The density function for each distribution and the corresponding notations for parameters are presented in Table 2.3.5. Two-sample Kolmogorov-Smirnov tests are carried out for the real data and the simulate data for 500 independent runs. This way, we can use the test statistics (denoted by H_M) and their corresponding p-values (denoted by P_M) as good indicators for our goodness-of-fit tests. Table 2.3.3 demonstrates the final results for all M's

[5]Note, method of moments here provides better fitting results for our data set when compare to Maximum likelihood estimation.

December				
Distributions	P_M	H_M	P_N	H_N
Log-Normal	0.8449	0	0.5493	0
Gamma	0.8299	0	0.4479	0
Normal	0.7156	0	0.2808	0
Weibull	0.4756	0	0.1642	0
Exponential	1.54E-19	1	1.92E-21	1
January				
Distributions	P_M	H_M	P_N	H_N
Normal	0.8191	0	0.3811	0
Weibull	0.7634	0	0.1119	0
Gamma	0.4507	0	0.5661	0
Log-Normal	0.2457	0	0.5948	0
Exponential	8.43E-16	1	4.28E-21	1
February				
Distributions	P_M	H_M	P_N	H_N
Normal	0.8946	0	0.7803	0
Gamma	0.7438	0	0.8038	0
Weibull	0.6986	0	0.5453	0
Log-Normal	0.6410	0	0.7904	0
Exponential	4.38E-20	1	5.12E-22	1

Table 2.3.3: Fitting results comparing to simulated distributions by Kolmogorov-Smirnov test.

	Gamma	Normal	Log-Normal	Weibull	Exponential
M	0.6748	0.8098	0.5772	0.6459	2.810E-16
N	0.6059	0.4807	0.6449	0.2738	2.239E-21

Table 2.3.4: 3-months average(each with 500 independent runs) Kolmogorov-Smirnov test p-values for each distribution of M and N

and N's from December to February. Note, H_M value equals to 0 if the test accepts the null hypothesis that the real data and the simulated data are from the same distribution at the 5% significance level, and 1 otherwise. In addition, from the p-values, one can tell generally Normal, Gamma, Weibull and Log-Normal distributions fit the data properly.

An average results of p-values is summarized in Table 2.3.4 for easy comparison. From the results in Tables 2.3.3 and 2.3.4, by comparing the p-values for both M and N, we propose 8 combinations of M and N in terms of CDFs in Table 2.3.6. Since the analytic solution is not available for most of the cases, we use numerical methods to generate the final results. Note that in order to keep the final results consistent

through out all three months, for each scenario, we keep the same distributions for M and N separately for all the three months.

All fitting results, with their estimated parameters for each M and N at each month, are presented in Table 2.3.7. The "Distribution" column indicates the CDFs that are used to fit the data, and "Real" in the first and second rows indicates the sample observation of the original data. We report the sample mean (μ) , standard deviation (σ). One can find that the original data has the same moments values as the estimation results for the Scenarios 1 and 2. [6]

2.3.1 Analytic solutions

In Figures 2.3.3 and 2.3.4, we draw the cumulative distribution functions of the loss variable $\mathbb{D}$ for December, January, and February given two different models of normal and exponential distributions for M, respectively. First of all, one can see that in both cases the probability of having no damage is rather high. While in the first model the probability of no damage is always more than 90%, in the second model it is more than 80%. Therefore, the second model is riskier. On the other hand, while the probability of having damages is always greater in January, the probability of having damages in December is always greater than February. A simple implication of this fact is that the prices of the insurance contracts in January is higher than December, and in December is higher than February. Finally, one can see the probability of damages with normal M is always lower than exponential M. The reason is that in the former case, there is always a possibility of good events (looking at Eq. (2.2.8)) while in the latter there is no such an event.

In Tables 2.3.8 and 2.3.9, we report the prices of two stop-loss policies for two different assumptions of a normal M and an exponential M, where analytical solutions can be found (Scenarios 1-2 in Table 2.3.6). Note, even though our tests on the historical data reject the scenarios where exponential distribution is used, for completeness, and the general application of the method on other data set, we still present the results here. In each table, we compute the prices of two layer stop-loss

[6]However, it can be instantly proven that exponential fittings are not optimal solutions as μ is not close to σ in all of our observations but, we still keep them here because first, the analytic solutions are available for them, and second, we believe in other real cases beyond this thesis, exponential and normal may result in better fitting results.

Name	Density	Parameters	Notes
Exponential	$f(x\|\mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}},\ x \geq 0$	μ	a special case of Gamma
Normal	$f(x\|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in \boldsymbol{R}$	μ, σ	
Log-Normal	$f(x\|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}},\ x > 0$	μ, σ	
Gamma	$f(x\|a,b) = \frac{1}{b^a\Gamma(a)} x^{a-1} e^{-\frac{x}{b}},\ x > 0$	a, b	$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx$
Weibull	$f(x\|A,B) = \frac{B}{A}(\frac{x}{A})^{B-1} e^{-(\frac{x}{A})^B},\ x \geq 0$	A, B	

Table 2.3.5: Distribution list used

	M's CDF	N's CDF	Duration	Results
Scenario 1	Exponential	Normal	3 months	Analytic
Scenario 2	Exponential	Exponential	3 months	Analytic
Scenario 3	Gamma	Gamma	3 months	Numerical
Scenario 4	Gamma	Log-Normal	3 months	Numerical
Scenario 5	Normal	Gamma	3 months	Numerical
Scenario 6	Normal	Log-Normal	3 months	Numerical
Scenario 7	Weibull	Gamma	3 months	Numerical
Scenario 8	Weibull	Log-Normal	3 months	Numerical

Table 2.3.6: Proposed Scenarios for M and N.

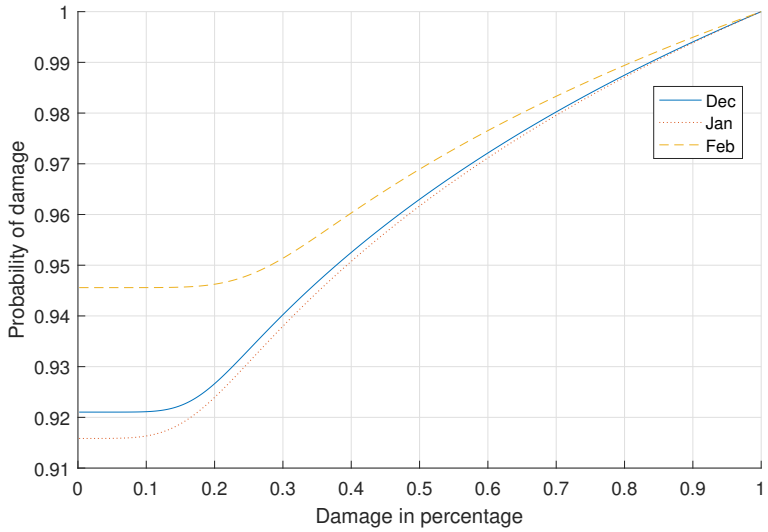


Figure 2.3.3: The CDF for December, January and February for exponential N and normal M, based on data from 1895 to 2016 in San Joaquin Drainage of California.

policies for $r = 0\%$ and $c = 90\%$, 95% and 100% coverage for December, January and February using three different risk premiums (distortion risk measures): VaR, CVaR and Wang's premium. Note that because of the shape of $\mathbb{D}$'s CDF, in all cases we do not have any changes in the premium prices for any r lower than 80%. That is why we consider $r = 0$ (i.e., a stop-loss insurance with lower retention level at 0 and higher retention level at c, both percentages). As one can see, the prices are fairly low for $c = 90\%$. The reason is that the probability of no damage, as discussed earlier, is pretty high. In addition, as discussed earlier again, since considering an exponential M will indicate higher risk of damages, for all cases the prices of insurance contracts for an exponential M tend to be higher than a normal M. We also have higher prices for CVaR comparing to Wang's premium and VaR. The reason is that CVaR puts a

Data	Distributions	December				January				February			
M	Real	μ	14.2421	σ	2.2174	μ	13.7451	σ	2.7944	μ	16.9123	σ	2.6476
N	Real	μ	9.7058	σ	1.5171	μ	9.4131	σ	1.4069	μ	10.4221	σ	1.3253
M	Exponential	μ	14.2421	-	-	μ	13.7451	-	-	μ	16.9123	-	-
N	Normal	μ	9.7058	σ	1.5171	μ	9.4131	σ	1.4069	μ	10.4221	σ	1.3253
M	Exponential	μ	14.2421	-	-	μ	13.7451	-	-	μ	16.9123	-	-
N	Exponential	μ	9.7058	-	-	μ	9.4131	-	-	μ	10.4221	-	-
M	Gamma	a	41.1275	b	0.3463	a	21.1401	b	0.6502	a	40.0105	b	0.4227
N	Gamma	a	42.6233	b	0.2277	a	46.3659	b	0.2030	a	61.7641	b	0.1687
M	Gamma	a	41.1275	b	0.3463	a	21.1401	b	0.6502	a	40.0105	b	0.4227
N	Log-Normal	μ	2.2609	σ	0.1533	μ	2.2313	σ	0.1472	μ	2.3358	σ	0.1285
M	Normal	μ	14.2421	σ	2.2174	μ	13.7451	σ	2.7944	μ	16.9123	σ	2.6476
N	Gamma	a	42.6233	b	0.2277	a	46.3659	b	0.2030	a	61.7641	b	0.1687
M	Normal	μ	14.2421	σ	2.2174	μ	13.7451	σ	2.7944	μ	16.9123	σ	2.6476
N	Log-Normal	μ	2.2609	σ	0.1533	μ	2.2313	σ	0.1472	μ	2.3358	σ	0.1285
M	Weibull	A	15.1964	B	6.9587	A	14.8483	B	5.5415	A	18.0402	B	6.8901
N	Gamma	a	42.6233	b	0.2277	a	46.3659	b	0.2030	a	61.7641	b	0.1687
M	Weibull	A	15.1964	B	6.9587	A	14.8483	B	5.5415	A	18.0402	B	6.8901
N	Log-Normal	μ	2.2609	σ	0.1533	μ	2.2313	σ	0.1472	μ	2.3358	σ	0.1285

Table 2.3.7: Fitting results for all scenarios

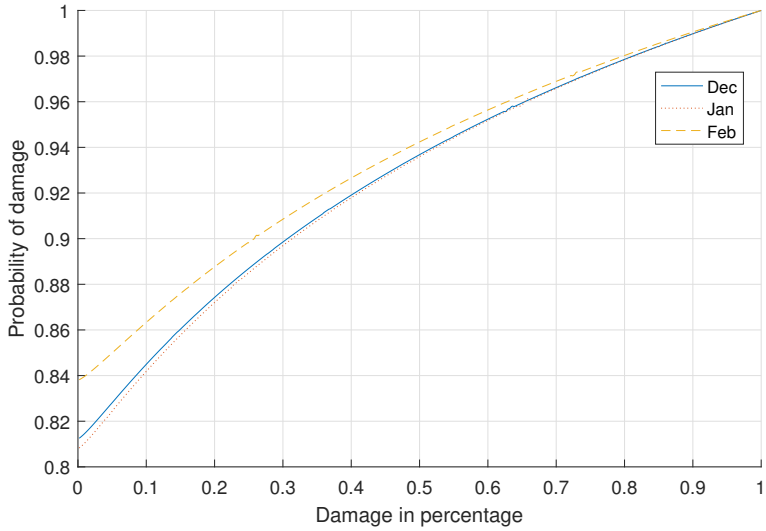


Figure 2.3.4: The CDF for December, January and February for exponential N and M, based on data from 1895 to 2016 in San Joaquin Drainage of California.

lot of weight on extreme events comparing to the two other premium rules.

2.3.2 Non analytic solution

Further to the previous two examples, we consider several more realistic scenarios (Scenarios 3-8 in Table 2.3.6) where numerical solutions are used to solve Eq.(2.2.5). Besides all similarities to our previous examples, we realize that in these cases, better model fittings lead to significant lower premiums. For instance, we witness more 0 premiums for the 95% VaR measure as shown in the first results column in the Table 2.3.10 to Table 2.3.11.

In Figure 2.3.5 and Table 2.3.10, the premiums are calculated where the monthly data of M and N are fitted into Gamma distributions. Based on the test statistics in Table 2.3.4, this is one of the preferable fitting configurations for our data set statistically. Compare to the analytic solutions, significant prices drop can be observed in all aspects from different measures to different thresholds. In particular, we see 0 prices in the 95% VaR measure for all three months. This means the underlying event is so extreme, that it is very unlikely to be observed if no more consideration is taken into the tail events. And such a solution is consistent for different thresholds and different measures.

		VaR			CVaR			Wang		
		95%	99%	99.5%	95%	99%	99.5%	$\beta = 1$	$\beta = 2$	$\beta = 3$
c=0.9	Dec.	0.3764	0.8364	0.9000	0.6456	0.8869	0.9000	0.2059	0.5251	0.7892
	Jan.	0.3927	0.8418	0.9000	0.6547	0.8884	0.9000	0.2111	0.5311	0.7921
	Feb.	0.2800	0.8091	0.9000	0.5952	0.8770	0.9000	0.1773	0.4901	0.7722
c=0.95	Dec.	0.3764	0.8364	0.9145	0.6499	0.9085	0.9419	0.2085	0.5382	0.8209
	Jan.	0.3927	0.8418	0.9164	0.6592	0.9107	0.9427	0.2137	0.5444	0.8240
	Feb.	0.2800	0.8091	0.9000	0.5989	0.8954	0.9366	0.1796	0.5023	0.8029
c=1	Dec	0.3764	0.8364	0.9145	0.6514	0.9157	0.9564	0.2096	0.5458	0.8447
	Jan.	0.3927	0.8418	0.9164	0.6607	0.9181	0.9576	0.2149	0.5521	0.8479
	Feb.	0.2800	0.8091	0.9000	0.6001	0.9015	0.9489	0.1806	0.5093	0.8257

Table 2.3.8: Prices of stop-loss policies for retention level 90%, 95% and 100% coverage for December, January and February using exponential N and normal M for three different risk premiums: VaR, CVaR and Wang premium.

		VaR			CVaR			Wang		
		95%	99%	995%	95%	99%	99.5%	$\beta = 1$	$\beta = 2$	$\beta = 3$
c=0.9	Dec.	0.5836	0.9000	0.9000	0.7666	0.9000	0.9000	0.2898	0.6130	0.8276
	Jan.	0.5873	0.9000	0.9000	0.7685	0.9000	0.9000	0.2923	0.6152	0.8283
	Feb.	0.5509	0.8927	0.9000	0.7503	0.8998	0.9000	0.2729	0.5976	0.8217
c=0.95	Dec.	0.5836	0.9018	0.9491	0.7740	0.9370	0.9491	0.2935	0.6293	0.8625
	Jan.	0.5873	0.9018	0.9491	0.7760	0.9373	0.9491	0.2961	0.6316	0.8634
	Feb.	0.5509	0.8927	0.9455	0.7571	0.9340	0.9489	0.2764	0.6134	0.8562
c=1	Dec	0.5836	0.9018	0.9491	0.7764	0.9494	0.9739	0.2952	0.6390	0.8895
	Jan.	0.5873	0.9018	0.9491	0.7785	0.9498	0.9741	0.2978	0.6413	0.8904
	Feb.	0.5509	0.8927	0.9455	0.7594	0.9455	0.9719	0.2780	0.6228	0.8827

Table 2.3.9: Prices of stop-loss policies for retention level 90%, 95% and 100% coverage for December, January and February using exponential N and M for three different risk premiums: VaR, CVaR and Wang premium.

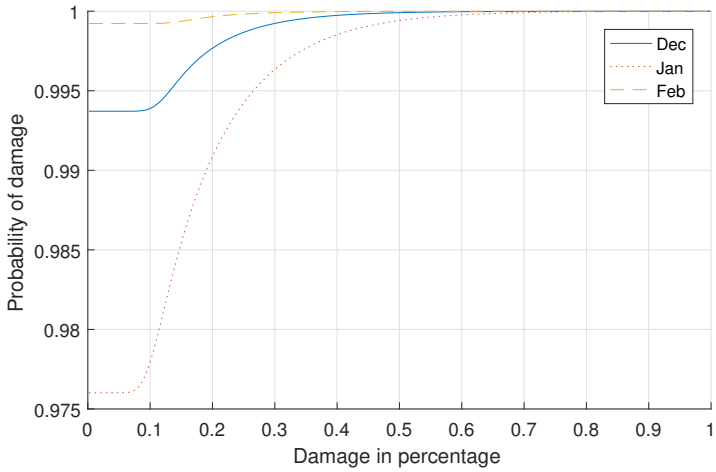


Figure 2.3.5: The commutative distribution function for December, January and February for Gamma N and M, based on data from 1895 to 2016 in San Joaquin Drainage of California.

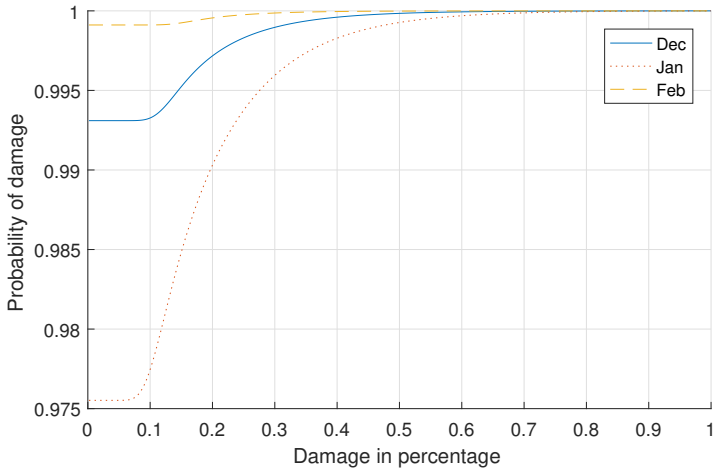


Figure 2.3.6: The commutative distribution function for December, January and February for Log-Normal N and Gamma M, based on data from 1895 to 2016 in San Joaquin Drainage of California.

		VaR			CVaR			Wang		
		95%	99%	99.5%	95%	99%	99.5%	$\beta = 1$	$\beta = 2$	$\beta = 3$
$C = 0.9$	Dec.	0.0000	0.0000	0.1309	0.0250	0.1252	0.2212	0.0159	0.0936	0.2931
	Jan.	0.0000	0.1891	0.2655	0.0969	0.2982	0.3738	0.0414	0.1754	0.4274
	Feb.	0.0000	0.0000	0.0000	0.0033	0.0164	0.0328	0.0036	0.0345	0.1583
$C = 0.95$	Dec.	0.0000	0.0000	0.1309	0.0251	0.1253	0.2212	0.0159	0.0937	0.2947
	Jan.	0.0000	0.1891	0.2655	0.0969	0.2982	0.3739	0.0414	0.1758	0.4314
	Feb.	0.0000	0.0000	0.0000	0.0033	0.0164	0.0328	0.0036	0.0345	0.1587
$C = 1$	Dec.	0.0000	0.0000	0.1309	0.0251	0.1253	0.2212	0.0159	0.0937	0.2955
	Jan.	0.0000	0.1891	0.2655	0.0969	0.2983	0.3739	0.0414	0.1760	0.4336
	Feb.	0.0000	0.0000	0.0000	0.0033	0.0164	0.0328	0.0036	0.0345	0.1588

Table 2.3.10: Prices of stop-loss policies for retention level 90%, 95% and 100% coverage for December, January and February using Gamma N and M for three different risk premiums: VaR, CVaR and Wang premium.

		VaR			CVaR			Wang		
		95%	99%	99.5%	95%	99%	99.5%	$\beta = 1$	$\beta = 2$	$\beta = 3$
$C = 0.9$	Dec.	0.0000	0.0000	0.1436	0.0290	0.1449	0.2434	0.0180	0.1046	0.3223
	Jan.	0.0000	0.1964	0.2745	0.1014	0.3104	0.3903	0.0433	0.1831	0.4437
	Feb.	0.0000	0.0000	0.0000	0.0039	0.0197	0.0395	0.0042	0.0393	0.1765
$C = 0.95$	Dec.	0.0000	0.0000	0.1436	0.0290	0.1449	0.2434	0.0180	0.1048	0.3248
	Jan.	0.0000	0.1964	0.2745	0.1015	0.3104	0.3904	0.0433	0.1837	0.4487
	Feb.	0.0000	0.0000	0.0000	0.0039	0.0197	0.0395	0.0042	0.0393	0.1771
$C = 1$	Dec.	0.0000	0.0000	0.1436	0.0290	0.1449	0.2434	0.0180	0.1049	0.3262
	Jan.	0.0000	0.1964	0.2745	0.1015	0.3104	0.3904	0.0433	0.1840	0.4515
	Feb.	0.0000	0.0000	0.0000	0.0039	0.0197	0.0395	0.0042	0.0394	0.1774

Table 2.3.11: Prices of stop-loss policies for retention level 90%, 95% and 100% coverage for December, January and February using Log-Normal N and Gamma M for three different risk premiums: VaR, CVaR and Wang premium.

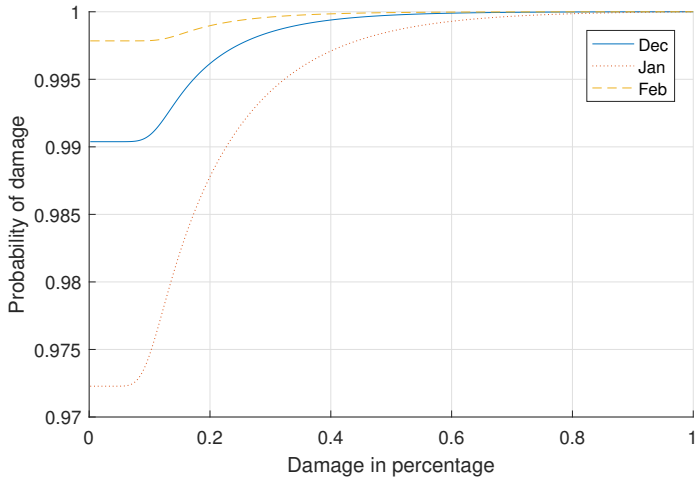


Figure 2.3.7: The commutative distribution function for December, January and February for Gamma N and Normal M, based on data from 1895 to 2016 in San Joaquin Drainage of California.

The same observations can be found in Figure 2.3.6 and Table 2.3.11 where the monthly data M and N are fitted in Gamma distribution and Log-Normal distribution, separately. One can find the premiums, are again, significantly lower than the analytical cases due to the goodness-of-fit. As it was expected, under the CVaR measure, all coverage settings have larger premiums when compare to the corresponding VaR measures, except the 0 prices cases. The same observation applies to Wang's measure when compare to VaR and CVaR.

Again in Figures 2.3.7 and 2.3.12, an increase of prices when α increase and prices are usually higher in January are consistent with previous results. Here, M and N are fitted into Gamma and Normal distributions, which are also adequate fittings for our data set when referencing Table 2.3.4. The overall prices are higher than the previous two cases, but still significantly smaller than the analytical solutions.

To conclude, Figure 2.3.5 to Figure 2.3.10 and Table 2.3.10 to Table 2.3.15 demonstrate how our pricing mechanism performs under Scenarios 3-8 as described in Table 2.3.6. Comparing to the analytical solutions, prices are lower in each configurations and even equals to 0 in some cases. This indicates the importance of fitting in the proposed model and how sensitive the tail event will affects the premiums - as one can see, the percentage change of $\alpha's$ is not linear to the corresponding change ratios of the premiums, and the later increases much faster. Another general observation through-

		VaR			CVaR			Wang		
		95%	99%	99.5%	95%	99%	99.5%	$\beta = 1$	$\beta = 2$	$\beta = 3$
$C = 0.9$	Dec.	0.0000	0.0000	0.1727	0.0400	0.2000	0.2787	0.0227	0.1224	0.3549
	Jan.	0.0000	0.2255	0.3218	0.1244	0.3644	0.4604	0.0518	0.2134	0.4971
	Feb.	0.0000	0.0000	0.0000	0.0097	0.0486	0.0972	0.0084	0.0646	0.2472
$C = 0.95$	Dec.	0.0000	0.0000	0.1727	0.0400	0.2000	0.2787	0.0228	0.1227	0.3580
	Jan.	0.0000	0.2255	0.3218	0.1245	0.3645	0.4607	0.0518	0.2146	0.5050
	Feb.	0.0000	0.0000	0.0000	0.0097	0.0486	0.0972	0.0084	0.0647	0.2489
$C = 1$	Dec.	0.0000	0.0000	0.1727	0.0400	0.2000	0.2787	0.0228	0.1228	0.3597
	Jan.	0.0000	0.2255	0.3218	0.1245	0.3646	0.4608	0.0519	0.2151	0.5097
	Feb.	0.0000	0.0000	0.0000	0.0097	0.0486	0.0972	0.0084	0.0647	0.2498

Table 2.3.12: Prices of stop-loss policies for retention level 90%, 95% and 100% coverage for December, January and February using Gamma N and Normal M for three different risk premiums: VaR, CVaR and Wang premium.

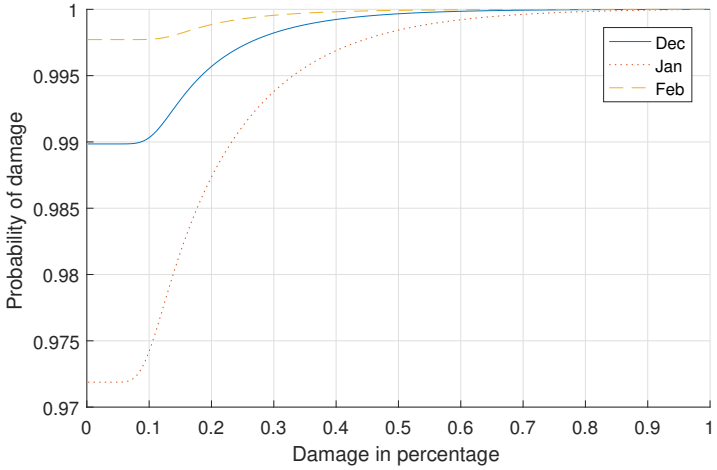


Figure 2.3.8: The commutative distribution function for December, January and February. for Log-Normal N and Normal M, based on data from 1895 to 2016 in San Joaquin Drainage of California.

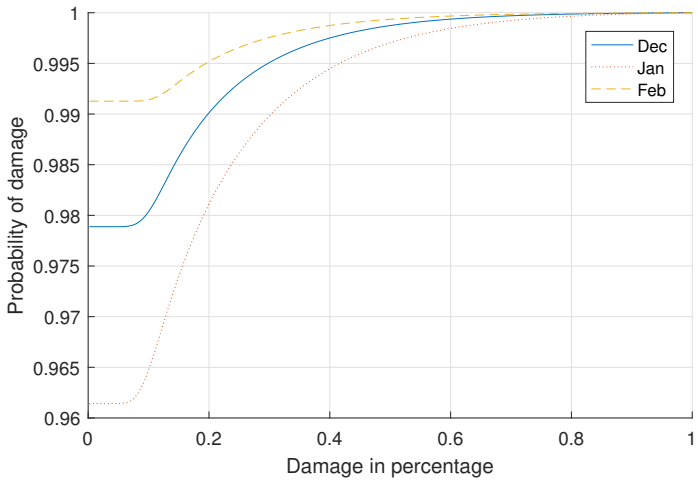


Figure 2.3.9: The commutative distribution function for December, January and February with Gamma N and Weibull M, based on data from 1895 to 2016 in San Joaquin Drainage of California.

		VaR			CVaR			Wang		
		95%	99%	99.5%	95%	99%	99.5%	$\beta = 1$	$\beta = 2$	$\beta = 3$
$C = 0.9$	Dec.	0.0000	0.0855	0.1836	0.0438	0.2176	0.2970	0.0246	0.1308	0.3740
	Jan.	0.0000	0.2309	0.3309	0.1281	0.3731	0.4713	0.0531	0.2180	0.5045
	Feb.	0.0000	0.0000	0.0000	0.0106	0.0529	0.1057	0.0090	0.0683	0.2581
$C = 0.95$	Dec.	0.0000	0.0855	0.1836	0.0438	0.2177	0.2971	0.0246	0.1312	0.3779
	Jan.	0.0000	0.2309	0.3309	0.1282	0.3733	0.4717	0.0532	0.2193	0.5129
	Feb.	0.0000	0.0000	0.0000	0.0106	0.0529	0.1057	0.0090	0.0685	0.2601
$C = 1$	Dec.	0.0000	0.0855	0.1836	0.0438	0.2177	0.2971	0.0246	0.1314	0.3802
	Jan.	0.0000	0.2309	0.3309	0.1282	0.3734	0.4718	0.0532	0.2199	0.5180
	Feb.	0.0000	0.0000	0.0000	0.0106	0.0529	0.1057	0.0090	0.0686	0.2611

Table 2.3.13: Prices of stop-loss policies for retention level 90%, 95% and 100% coverage for December, January and February using Log-Normal N and Normal M for three different risk premiums: VaR, CVaR and Wang premium. (Note the value and the settings for the number in bold here, later, a robustness study will be based on this configuration)

		VaR			CVaR			Wang		
		95%	99%	99.5%	95%	99%	99.5%	$\beta = 1$	$\beta = 2$	$\beta = 3$
$C = 0.9$	Dec.	0.0000	0.1982	0.2964	0.0987	0.3404	0.4398	0.0448	0.1979	0.4815
	Jan.	0.0000	0.3018	0.4145	0.1882	0.4571	0.5620	0.0703	0.2628	0.5603
	Feb.	0.0000	0.0000	0.1945	0.0450	0.2251	0.3401	0.0266	0.1454	0.4131
$C = 0.95$	Dec.	0.0000	0.1982	0.2964	0.0987	0.3406	0.4401	0.0449	0.1990	0.4891
	Jan.	0.0000	0.3018	0.4145	0.1883	0.4575	0.5629	0.0705	0.2647	0.5711
	Feb.	0.0000	0.0000	0.1945	0.0450	0.2252	0.3403	0.0266	0.1462	0.4190
$C = 1$	Dec.	0.0000	0.1982	0.2964	0.0987	0.3406	0.4402	0.0449	0.1995	0.4936
	Jan.	0.0000	0.3018	0.4145	0.1883	0.4577	0.5631	0.0705	0.2656	0.5777
	Feb.	0.0000	0.0000	0.1945	0.0450	0.2252	0.3403	0.0266	0.1465	0.4225

Table 2.3.14: Prices of stop-loss policies for retention level 90%, 95% and 100% coverage for December, January and February using Gamma N and Weibull M for three different risk premiums: VaR, CVaR and Wang premium.

		VaR			CVaR			Wang		
		95%	99%	99.5%	95%	99%	99.5%	$\beta = 1$	$\beta = 2$	$\beta = 3$
$C = 0.9$	Dec.	0.0000	0.2036	0.3055	0.1022	0.3497	0.4515	0.0462	0.2028	0.4896
	Jan.	0.0000	0.3055	0.4218	0.1914	0.4635	0.5697	0.0714	0.2660	0.5648
	Feb.	0.0000	0.0000	0.1982	0.0463	0.2315	0.3469	0.0272	0.1481	0.4183
$C = 0.95$	Dec.	0.0000	0.2036	0.3055	0.1022	0.3499	0.4519	0.0463	0.2040	0.4978
	Jan.	0.0000	0.3055	0.4218	0.1915	0.4639	0.5706	0.0715	0.2680	0.5760
	Feb.	0.0000	0.0000	0.1982	0.0463	0.2316	0.3471	0.0272	0.1489	0.4246
$C = 1$	Dec.	0.0000	0.2036	0.3055	0.1022	0.3499	0.4520	0.0463	0.2046	0.5027
	Jan.	0.0000	0.3055	0.4218	0.1915	0.4640	0.5708	0.0716	0.2689	0.5829
	Feb.	0.0000	0.0000	0.1982	0.0463	0.2316	0.3471	0.0273	0.1493	0.4283

Table 2.3.15: Prices of stop-loss policies for retention level 90%, 95% and 100% coverage for December, January and February using Log-Normal N and Weibull M for three different risk premiums: VaR, CVaR and Wang premium.

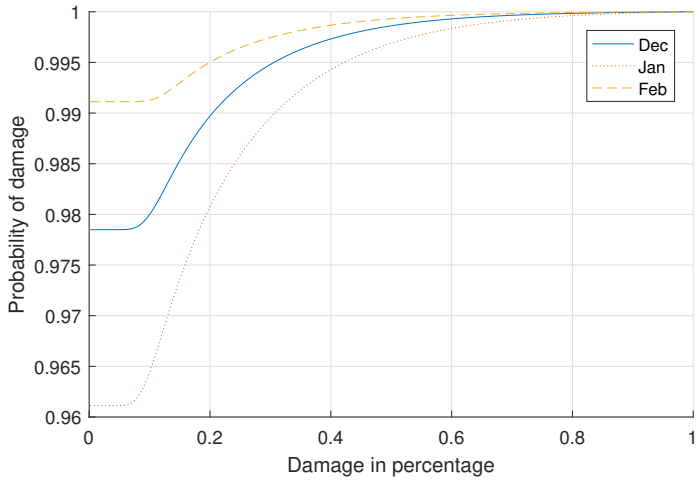


Figure 2.3.10: The commutative distribution function for December, January and February for Log-Normal N and Weibull M, based on data from 1895 to 2016 in San Joaquin Drainage of California.

out all solutions is that January is always the month that has the largest premiums. This is true for different risk measure, different $\alpha's$ and different thresholds.

2.3.3 Summary of data discussions

We summarize the results of our study by providing Table 2.3.16, in which the maximum and the minimum premium prices for all the scenarios above are demonstrated. Since all the scenarios are only separated by different CDFs fittings to M and N, one can see Scenario 3 gives the lowest premiums, and Scenario 2 gives the largest in terms of average premium over all risk measures and coverage. Note, if we are only concerned with the average p-values from the Kolmogorov-Smirnov test, in all of our simulations, Scenario 6 has the optimal fitting results for both M's and N's CDFs. But this combination does not yield the minimum premium price anywhere among the maximum nor average prices over all risk measures. Thus we believe an inaccurate fitting may result overpricing or under-pricing for insurance contacts.

2.3.4 Robustness analysis

It is rather intuitive to assume that a harvest with high resistance to a low temperature (i.e., high k) requires lower insurance premium. To test this theory and see how our method performs under different parameter settings for different types of agriculture

	M's CDF	N's CDF	Maximum	Average	Minimum
Scenario 1	Exponential	Normal	0.9576	0.6769	0.1773
Scenario 2	Exponential	Exponential	0.9741	0.7566	0.2729
Scenario 3	Gamma	Gamma	0.4336	0.1121	0.0000
Scenario 4	Gamma	Log-Normal	0.4515	0.1200	0.0000
Scenario 5	Normal	Gamma	0.5097	0.1459	0.0000
Scenario 6	Normal	Log-Normal	0.5180	0.1550	0.0000
Scenario 7	Weibull	Gamma	0.5777	0.2345	0.0000
Scenario 8	Weibull	Log-Normal	0.5829	0.2389	0.0000

Table 2.3.16: Summary of extreme prices for all Scenarios. (Note, scenario 6 in bold here is the scenario with best fitting results for M and N according to 2.3.4)

products, in the following section, we conduct a robustness analysis. Table 2.3.17 summaries 25 settings. We assume that T_t and T_c range from $17°F$ to $23°F$, and $25°F$ to $31°F$, respectively and can resist from $k = 3$ hours to $k = 5$ hours of low temperature with an interval of 0.5 hour. To simplify the comparison and focus on the price-product relation, only January's data is used. Beside that, we also set the risk measure as Wang's premium with $\beta = 2$ all the time, and the retention levels for stop-loss insurance are set at 0 to 1 which means a full coverage for all 25 cases.

In the centre of Table 2.3.17, we have the same setting (number is Bold) as in Scenario 6 for Wang's premium with $c = 1$, $\beta = 2$ and $\eta = 1.5$, which gives the insurance premium of 0.2199. As we move around the centre, the premium becomes higher as k decrease, meaning that the resistance time has reduced. The premium also decreases as T_t and T_c become smaller, which means that the lower temperature one product can resist to, the cheaper the premium will be. In Figure 2.3.11, we simulate a wider range of scenarios, where $T_t = 15, 16, 17, ..., 28$, $T_c = 19, 20, 21, ..., 32$ and $k = 1, 1.5, 2, ..., 10$, to test the stability of the proposed methods empirically, and demonstrate the flexibility for using such models.

2.4 Interim Conclusion

In this chapter, we provided a framework for pricing frost insurances. The reason for doing this research is because frost is a type of natural disaster with adverse effects on agriculture products with very little damage report. We want to create a new method that can efficiently and accurately price frost insurance based on the relevant data

T_t	17	20	20	20	23
T_c	25	25	28	31	31
$k = 3$	0.0648	0.1106	0.2244	0.3469	0.4854
$k = 3.5$	0.0631	0.1070	0.2225	0.3463	0.4845
$k = 4$	0.0609	0.1025	0.2199	0.3456	0.4833
$k = 4.5$	0.0584	0.0971	0.2166	0.3446	0.4816
$k = 5$	0.0554	0.0909	0.2124	0.3432	0.4793

Table 2.3.17: Sensitivity test summary of insurance prices. (Note, the number in bold at the centre is with same configuration as we demonstrate in Table 2.3.13)

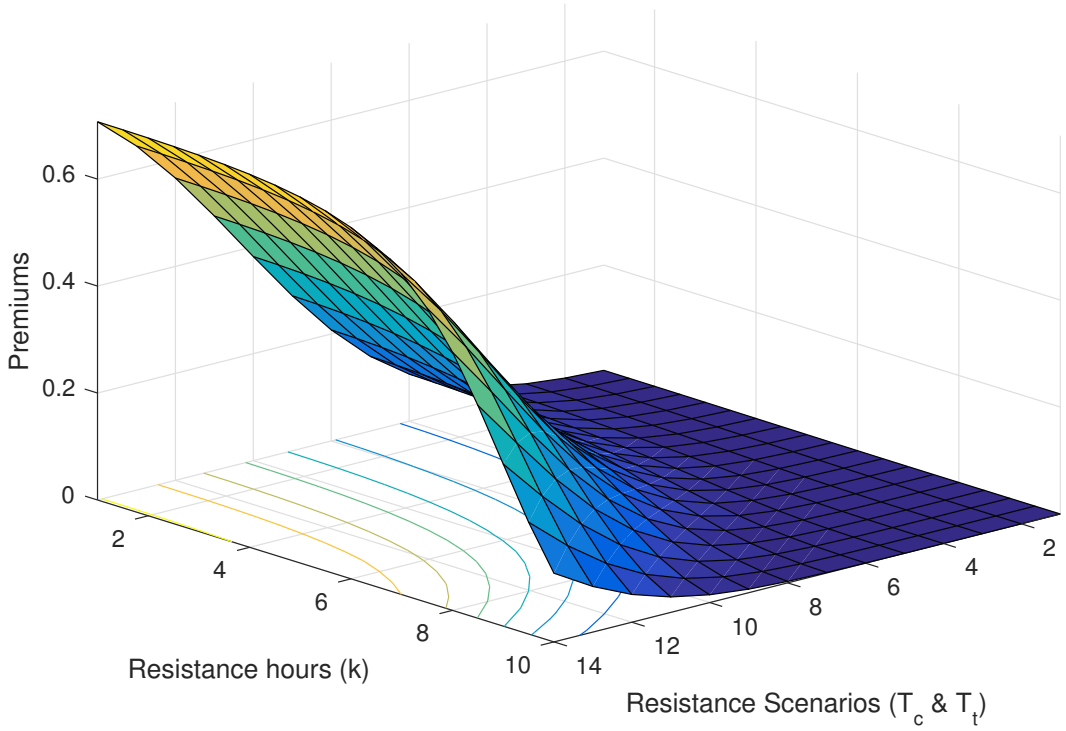


Figure 2.3.11: 3D plots of insurance prices for sensitivity test for different resistance hours (k) and scenarios ($T_t,\ T_c$).

one can easily collect. Also, the work we have done in Chapter 2 answers Question 3 in great detail.

The approach we took is to first introduce a frost-loss model based on temperature that can address the non-linearity of the frost losses with respect to temperature. The link between frost loss and temperature is important, as stated before, given that frost damage data are hard to find while temperature data are complete almost anywhere. Then, we introduced a pricing method where the risk is measured by a general distortion risk measure. We show that how distortion risk premiums can properly price a large family of insurance contracts with no risk of moral hazard, including multiple-layer stop-loss policies.

More specifically, we found a model for the frost-loss damages. However, in comparison with existing studies, such as [42] and [99], we decided to address two further important issues. The first issue is that the damage to the harvest is nonlinear (convex) in terms of the temperature drops. The existing methods usually simplify this mechanism by setting a Bernoulli distributed loss variable for frost damage; see [99]. Second, we realise from the literature that low temperatures would not cause any damage unless the harvest experiences at least several hours of the low temperature. Thus, following the approach in [109], we introduced a variable D in Eq.(2.2.2), where the percentage of frost damage, as well as temperature changes, are all modelled in a single variable. After deriving the solutions for Eq.(2.2.2), we further address the second complication where we incorporate the harvest resistance to low temperatures for a few hours in Ep.(2.2.4). By making some relevant assumptions, we found that the probability of having no damage is rather high. This is particularly true in December and February over the course of the winter.

Finally, we tested our models by pricing stop-loss insurance policies based on data sets from San Joaquin Drainage County in California, using different risk premiums rules: Value at Risk, Conditional Value at Risk and Wang's premium. As we realised earlier from the distribution for losses, $\mathbb{D}$, in all scenarios with different premium rules, the premium for January is the largest, more than December and February. The result is consistent with experience and, again, provides quantifiable measurements for pricing frost insurance and the difference in premiums under each risk measure.

Chapter 3

Hurst exponent estimation

In order to better modelling the long-memory property, which usually exhibits by commodity data, we introduce a recursive method for estimating the Hurst exponent of a time series in this Chapter. Such a method relies on mathematical relations between the Hurst exponent and the fractal parameter of an ARFIMA model. Our method estimates the Hurst exponent by applying an auto-regressive filter to the data repeatedly until its value converges. Following similar studies in this ares like [107] and [79], we first apply this method on simulated data and observe its stability and convergence. Then we estimate the Hurst exponents of commodity prices and observe how Hurst exponent values are different for agricultural and non-agricultural commodity prices. In addition, we find the optimal ARFIMA models for each commodity, which is useful for simulate scenarios for risk management.

3.1 Spectral density and Hurst coefficient

In economics and finance, long memory-ness was first observed in [39] where the author observed that economic and financial historical data typically exhibit some distinct low frequent non-periodic cyclical patterns. Simultaneously, in [47] and [41], the ARFIMA model were introduced which greatly improved the applicability of LRD studies in statistical practices. Since the 1990s, with the new sources of financial data, there has been a surge of interest in analysing financial data with LRD. The author in [23] discussed the relevance between LRD and financial modelling, as well as their relation with the basic principles of financial theory and possible economic explanations for their presence in financial time series. On the conceptual and empirical level,

the author concluded that statistical analysis alone is not likely to provide a definite answer for the presence or absence of LRD. In [40], it is suggested that long memory in an economic time series can be due to the aggregation of a cross section of time series with different persistence levels. In [57], the author discussed the evolutionary models' impact on LRD, and Kirman in [54] explored the mechanism whereby switching of agents' trading behaviours between two or more strategies could be a cause for LRD in financial data. Finally, Liu in [60] argued that the presence of a Markovian regime switching mechanism can lead to volatility clustering but not necessarily LRD.

Let us formally introduce the Long-Range Dependence.

Definition 12. Let Y_t be a second-order stationary process with auto-covariance function $\gamma_Y(k)$ $(k \in \mathbb{Z})$ and spectral density

$$f_Y(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_Y(k) e^{-ik\lambda}, \lambda \in [-\pi, \pi].$$

Introduce the symmetric function $L_f(\lambda) = \frac{f_Y(\lambda)}{|\lambda|^{-2d}}$ and assume that it is slowly varying at zero, i.e. $\forall a > 0, \frac{L_f(at)}{L_f(t)} \to 1$ as $t \to 0$. Then Y_t exhibits (linear)

(a) long-range dependence, if $d \in (0, \frac{1}{2})$

(b) intermediate dependence, if $d = 0$ and $\lim_{\lambda \to 0} L_f(\lambda) = \infty$;

(c) short-range dependence, if $d = 0$ and $\lim_{\lambda \to 0} L_f(\lambda) \in (0, \infty)$;

(d) anti-persistence, if $d \in (-\frac{1}{2}, 0)$.

We introduce a simple fractional series as follows

Definition 13. A series X_t is called simple fractional series if it has the spectral density $f_2(\lambda) = (\sigma^2/2\pi)|1 - e^{-i\lambda}|^{-2d} = (\sigma^2/2\pi)(2\sin(\lambda/2))^{-2d}$, for some $d \in (-1/2, 1/2)$.

Definition 14. Based on Definition 12, let us denote the short memory (short-range dependence) spectral density by $f_u(\lambda)$. Y_t is called a general integrated series if its spectral density is of the form $f_2(\lambda) f_u(\lambda)$ where $f_2(\lambda)$ is given in Definition 2.

In [41], the ARFIMA process is referred to a model of time series with long-range dependence. For a stationary time series, an $ARFIMA(p, d, q)$ model is an ARMA model where the innovations are fractional white noise, which can be rewritten in lag operator notation as:

Definition 15. Let Y_t be a stationary process with $\epsilon_t \sim iid\ N(0, \sigma_\epsilon^2)$ such that

$$\Phi(B)(1-B)^d(Y_t - \mu) = \Theta(B)\varepsilon_t$$

where d is a fractional integration parameter that is allowed to take non-integer values, B is the lag operator ($B^n Y_t = Y_{t-n}, n = 1, 2, 3...$), $\Phi(B) = (1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)$ specifies the AR lag polynomial, and $\Theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)$ specifies the MA lag polynomial.

It is clear that an $ARFIMA(p, d, q)$ is a general integrated series where $f_2(\lambda) = (\sigma^2/2\pi)|1 - e^{-i\lambda}|^{-2d} = (\sigma^2/2\pi)(2\sin(\lambda/2))^{-2d}$ and f_u is associated with the $ARIMA$ part.

The properties of an $ARFIMA$ process Y_t depend on the value of the fractional integration parameter d. However, many researchers such as [97] and [15] suggest the estimations based on heuristic methods are more robust, and this is the reason why in this thesis we focus on the heuristic approach to detect LRD.

The following fundamental proposition, relating Hurst exponent and fractional coefficient d of an $ARFIMA$ process, was proven in [38].

Theorem 1. *If Y_t is an ARFIMA process with parameter $d \in (-0.5, 0.5)$, then it is also a stationary process with Hurst exponent $H = d + 0.5$.*

In empirical studies, it is well known that the estimation of $ARFIMA$ model is not stable due to the ill-behaved likelihood function, see [80, 81, 83, 87, 96, 95, 82] for details. Here, we use Theorem 1 to construct a convergence algorithm to fit the data. The algorithm estimates an initial Hurst exponent first, then several loops follow-by to de-fractionalise underlying data until the most stable Hurst exponent is found. In the final step, the Hurst exponent is transferred to the fractional parameter to fit the selected $ARFIMA$ model. We first test this recursive approach to estimate the Hurst exponent and the $ARFIMA$ parameters for simulated data. Then, we will

apply the algorithm to a set of commodity price data with the matching Hurst index and optimal $ARFIMA$ models being reported.

Regarding the estimation methods for $ARFIMA$ processes, there are three main approaches: maximum likelihood methods (see [113], [100], [24]), semi-parametric methods (see [38], [86]) and heuristic methods (see [50], [45], [62]). The maximum likelihood estimator provides a consistent approach to estimate all parameters of interest simultaneously, but it usually generates unstable results with high computational costs due to the ill-behaved likelihood function. The other two approaches are also called two-step estimation methods, where, to fit an $ARFIMA(p, d, q)$ model, we first estimate the long memory parameter d and then we fit an $ARMA$ model to the data. More specifically, the Geweke & Porter-Hudak (or GPH) method [38] is based on the behaviour of the spectral density of the $ARFIMA$ process near frequency zero, but it shows a bias in the presence of strong autocorrelation in the $ARMA$ process. A modified version of the GPH estimator that uses a smoothed periodogram was introduced to reduce the bias by [86].

The exact log-likelihood function for a $ARFIMA(p, d, q)$ process y_t is

$$logL(d, \phi, \theta, \sigma_\epsilon^2) = -\frac{T}{2}log(2\pi) - \frac{1}{2}log|\Sigma(d, \phi, \theta, \sigma_\epsilon^2)| - \frac{1}{2}z'\Sigma(d, \phi, \theta, \sigma_\epsilon^2)^{-1}z.$$

where d is the fractional integration parameter, ϕ is the parameters for AR part, θ is the parameters for MA part, $z = y - \mu$, and Σ is the auto-covariance matrix of $y = (y_1, \cdots, y_T)$

$$\Sigma(d, \phi, \theta, \sigma_\epsilon^2) = V[y] = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \gamma_1 \\ \gamma_{T-1} & \cdots & \gamma_1 & \gamma_0 \end{pmatrix}.$$

Here, γ_i is the auto-covariance function of an $ARMA$ model with mean μ

$$\gamma_i = E[(y_t - \mu)(y_{t-i} - \mu)].$$

In order to show the instability of the standard maximum likelihood estimator approach and the problem of parameter miss-specification, we depict the likelihood function on an $ARFIMA(1, d, 0)$ process. Figure 3.1.1 shows the relationship between

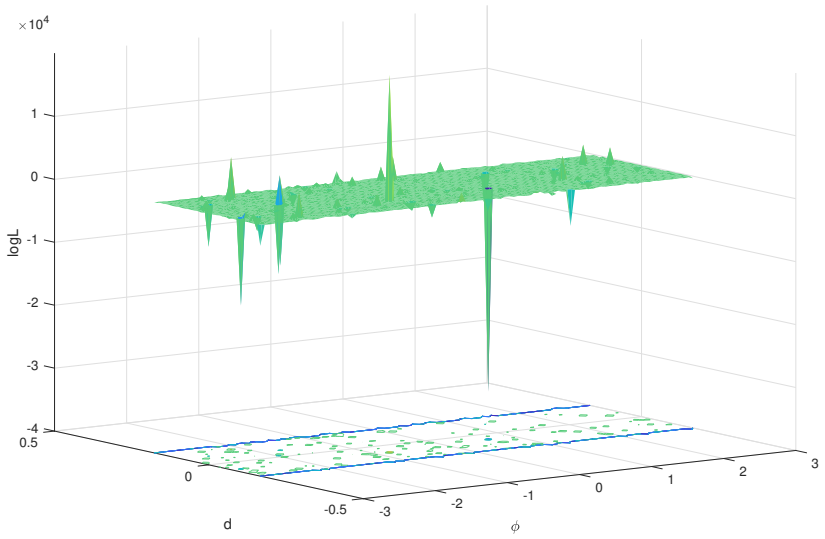


Figure 3.1.1: $logL$ vs ϕ,d

$logL$ and ϕ,d, fixing $\sigma_\epsilon^2 = 1$. Figure 3.1.2 shows the relationship between $logL$ and d,σ_ϵ^2, fixing $\phi = 0.5$. Figure 3.1.3 shows the relationship between $logL$ andϕ,σ_ϵ^2, fixing $d = 0.3$.

As one can see from the graphs, the ill-behaviour likelihood function makes it difficult to find the maximum points as it would be trapped in the local maximum points easily. Even though we find the maximum points based on appropriate initial values from the three groups separately, it is rather rare for their combinations (i.e., the combination of ϕ, d, σ_ϵ^2) to be maximum. Thus, a direct implementation of maximum likelihood estimator is not recommended.

Although time series models like $ARFIMA$ are developed fairly well, several researchers such as [15] and [97] suggest that estimating time series data based on heuristic methods are more robust. By 'heuristics' we mean specific techniques that are developed for the parameter estimation of Hurst components instead of the general likelihood method or Method of Moments.

In the following, we will see if heuristic approaches to time series estimation of the spectral density and Hurst coefficient can improve the estimations obtained by the $ARMA$ and $ARFIMA$ methods. To accomplish this, we present a new algorithm to determining H from an $ARFIMA$ model and compare the outcomes to the known approaches in the literature. We show that the algorithm improves the robustness of

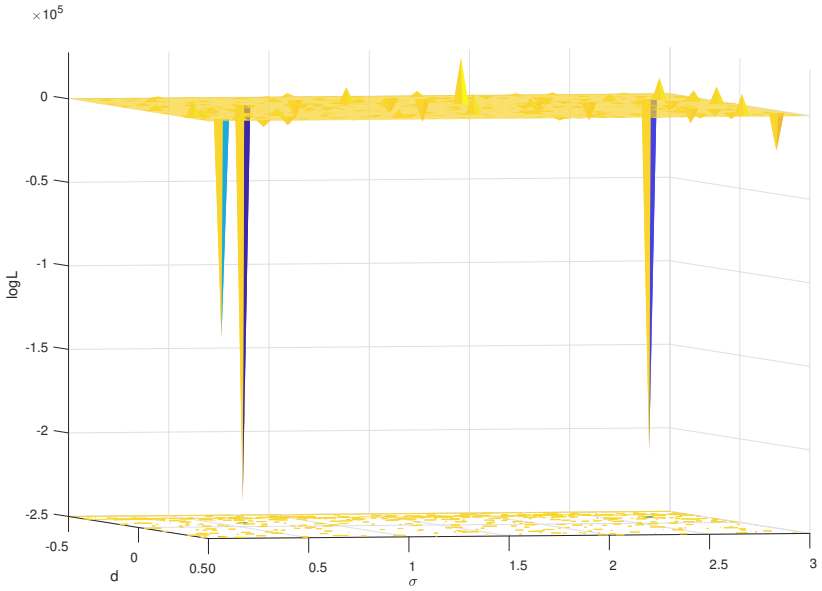


Figure 3.1.2: $logL$ vs d,σ_ϵ^2

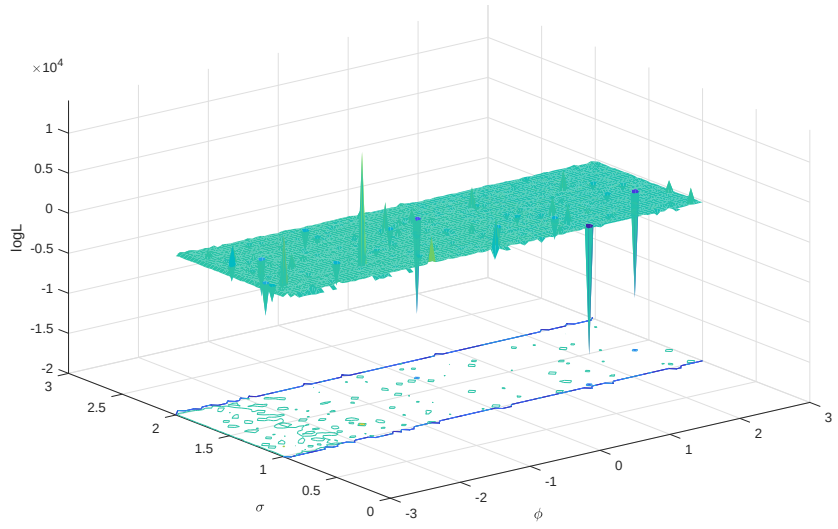


Figure 3.1.3: $logL$ vs ϕ,σ_ϵ^2

the estimates as measured by lower mean error and standard deviation. We do this experimentally using simulated data derived from a random number generator with known H. We then apply the technique to determine the H coefficients for a number of commodity futures prices and indices.

3.2 Methodologies on estimating the Hurst exponent

Due to the nature of our study, we have chosen methods for estimating the Hurst exponent based on stability and efficiency. Thus, three methods are used: Re-scaled Range Statistic Method (R/S in short), Aggregated Variance Method (AV in short) and a recursive implementation of the AV method (OT in short); see [82].

3.2.1 Existing methods estimate the Hurst exponent

Here, we discuss three estimation methods for the Hurst exponent: Range Statistic method, Aggregated Variance method and an extension of the Aggregated Variance Method we have implemented.

3.2.1.1 Range Statistic method

The Range Statistic method or the re-scaled range (or R/S) analysis, is one of the best-known technique for estimating H. It is able to distinguish a random series from a fractal series, irrespective of the distribution of the underlying series (Gaussian or non-Gaussian).

Given a sequence of n observations $(y_1, y_2, \cdots, y_n)$, $R = R(n)$ is the difference between the maximum and the minimum cumulative deviations of the observations y_t from its mean μ i.e.,

$$R(n) = \max_{1 \leq t \leq n} \left[\sum_{i=1}^{t} (y_i - \mu) \right] - \min_{1 \leq t \leq n} \left[\sum_{i=1}^{t} (y_i - \mu) \right].$$

Let $S = S(n)$ denote the standard deviation of the original time series i.e.,

$$S(n) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \mu)^2}.$$

The Hurst exponent H is a number that the following holds

$$E(R(n)/S(n)) = Cn^H$$

where C is a positive, finite constant independent of n. As a result, the Hurst exponent H can be calculated from

$$H = \frac{\log(R(n)/S(n))}{\log(n)}, \ 0 < H < 1.$$

Hence, using a simple regression, the estimate of H can be found by calculating the slope of $\log(R/S)$ against $\log(n)$. Based on our observation if the data, for R/S method, we chose a constant time span of 2^{10} as the observation rolling window size. This method is able to obtain a robust estimator even for the data followed by a heavy-tailed probability density function (see [68]). On the other hand, the exact distribution of the R/S statistic is difficult to determine, since it is usually biased and affected by the non-stationary property of the data. One possible way is discussed in [75]: that is to reduce the bias by ignoring the points on the extreme left and extreme right of the log-log plot; the former due to the influence of the short-term dependence structure and the latter due to only a few observations. However, another possible approach is applied in this thesis: Instead of using the slope of common logarithm (using base 10) of R/S against $\log(n)$ we are going to adjust it to the binary logarithm (using base 2), that is $H = \log_2(R/S)/\log_2(n)$. Note, doing so will reduce the data outliers for the linear regressions, thus allows us to estimate the slope with all the available points.

3.2.1.2 Aggregated Variance method

Generally speaking, the Aggregated Variance method (AV in short) introduced in [11] is the analysis of the variances of aggregated time series processes. Given a sequence of n observations $(y_1, y_2, \cdots, y_n)$, one property of a long-memory process is that the variance of the sample mean μ_n converges to zero slower than the rate n^{-1}, where n is the sample size. For a large sample size n, they have the following relationship (see [10]):

$$Var(\mu_n) \sim cn^{2H-2},$$

where $c > 0$. Using the AV method to estimate H, the underlying series is first divided into n/m blocks (or bins) of size m, and sample mean is computed for each block

$$\mu_m(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} y_i \,, \quad k = 1, 2, \cdots, n/m,$$

with sample variance

$$s^2(m) = (n/m - 1)^{-1} \sum_{k=1}^{n/m} [\mu_m(k) - \mu_n]^2.$$

Hence, plotting $\log s^2(m)$ against $\log(m)$ should yield points scattered with slope equal to $2H - 2$. The Hurst exponent H can then be estimated by evaluating the slope through regression, where, again, a first-order linear function is adopted. For the AV method, we chose a constant block size of 1000 for splitting the data.

Extended Aggregated Variance method

A slight improvement on the aggregated variance method is noted when we updated the block size continuously within the estimation process. We implemented such a method (OT in short) where, instead of using a fixed size block size for splitting the data, we dynamically adjusted it during the estimation process until the block size is smaller than 5.

3.2.2 Stability for current methods

In Table A.4.4, we generated synthetic time series with known $H = d + 0.5$ coefficients using $ARFIMA(0, d, 0)$ process and a first difference fraction Brownian motion (fBm). We then compared the H statistics, as discussed above, to measure the deviation from $d = H - 0.5$. We can then test the accuracy of the three chosen methods on estimating the Hurst exponent for simulated data at all Hurst ranges. For generating artificial $ARFIMA$ process, we chose the fast algorithm processes in [53] where the calculation speed (number of arithmetic operations) is improved from an order of T^2 to an order of $T \log(T)$, where T is the length of the time series.

Recall that, in an $ARFIMA$ process, the process is stationary if $-0.5 < d < 0.5$, and according to Theorem 1, $d = H - 0.5$, i.e. $0 < H < 1$. Thus, we

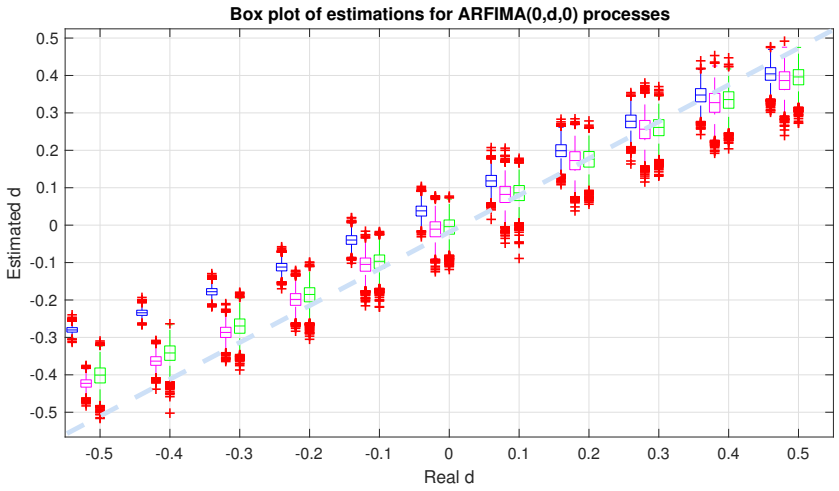


Figure 3.2.1: $ARFIMA(0, d, 0)$ process box-plot (Blue: RS, Magenta: AV, Green: OT)

simulate 10000 data points for each of the $AFRIMA(0, d, 0)$, $AFRIMA(0.5, d, 0)$, $AFRIMA(0, d, 0.3)$, $AFRIMA(0.5, d, 0.3)$ and fractional Brownian Motion processes with known fractional parameter d ranges from -0.5 to 0.5 separately (or $H = 0\ to\ 1$). The simulated procedure involves 5000 random draws for each of the 10000 observations from known H distributions. For each iteration, we considered 55 separate bins and computed the mean values in each bin. The box plot for the $ARFIMA(0, d, 0)$ process showing the means and range for each setting are shown in 3.2.1. Note, first, that the best fit is that with best projects along a 45 degree line from the origin; second, since the algorithm requires stationary data, first difference or log transfer should be applied to data if necessary. The calculated average mean error and average standard deviation in each of the five models are presented in Table A.4.4 for comparison. Generally speaking, the method with the smallest average error and average standard deviation is the most accurate one for all ranges of Hurst exponent estimations. We find that the AV and OT method perform better in terms of mean values. However, for a comprehensive study, we keep all three estimations in the following study and use the average of them as our initial Hurst exponent estimation.

In Figures 3.2.1 and 3.2.2, we visualise part of the simulated estimation results for three different methods in box-plots. Again, for the $ARFIMA(0, d, 0)$ and $ARFIMA(p, d, q)$ models, one can see all methods generally work well when $d \geqslant 0$, i.e. $H \geqslant 0.5$. However, we notice that the traditional heuristic approaches work bet-

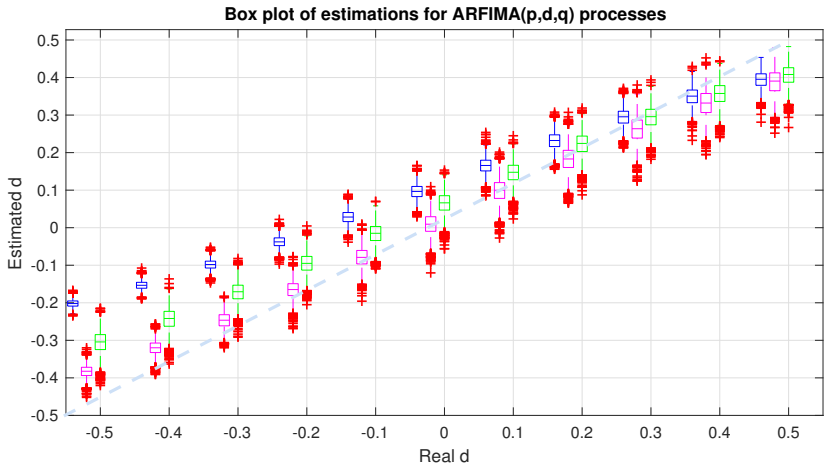


Figure 3.2.2: $ARFIMA(0.5, d, 0.2)$ process box-plot (Blue: RS, Magenta: AV, Green: OT)

ter with $ARFIMA(0, d, 0)$ model or the pure fractional noise. This can be explained by the short memory part of the time series which can act like a disruptor when the measure for long memory is applied. From this observation, one can remove the short memory part from a time series data, and only estimate the Hurst exponent for the long memory part. Thus, a more accurate fitting algorithm may be found for the measure of long memory.

3.2.3 A recursive approach

Adopting the idea from [82], where the authors used an algorithm on testing the stability of a Hurst estimation, we used a convergence procedure to estimate the parameters for an $ARFIMA$ model. The procedure goes as follows: we first estimate an initial Hurst exponent from the methods we mentioned above, then an infinite auto-regressive filter, $Y_t = (1-B)^d X_t = \sum_{s=0}^{\infty} b(s) B^s X_t$, is applied on the data X_t with the Hurst exponent satisfying $d = H - 0.5$. With the newly generated data Y_t, we adopt a traditional time series methodology (i.e. Box-Jenkins Method) to find an adequate $ARIMA$ model. After estimating the polynomials of the $ARIMA$ model, a new time series is generated by the filter: $\widehat{Y}_t = \dfrac{\widehat{\Phi}(B)}{\widehat{\Theta}(B)} X_t$ in order to find the "pure" $ARFIMA(0, d, 0)$ process. Finally, a recursive method is applied to estimate Hurst exponent until the Hurst exponent converges. We provide the following list for the details of each step in our algorithm:

1. De-seasonalise and adjust for inflation if needed.

2. Check data's stationary property and generate the initial time series data X_t.

   (a) Carry out Augmented Dickey-Fuller (ADF) test.

   (b) Take the increment, log difference or return ratio if needed.

   (c) Adjust observed data and generate the initial stationary time series X_t.

3. Carry out initial calculation of the Hurst parameter for X_t.

   (a) Use re-scaled Range Statistic Method (R/S procedure).

   (b) Use aggregated Variance Method.

   (c) Use recursive implementation of the AV method.

   (d) Use the average of (a), (b) and (c) as our initial estimation for Hurst.

4. Use $H = d + 0.5$ to find the initial value of d; the fractional parameter for the $ARFIMA$ process. (See details and proofs in [38]. Note, H is from d, but one cannot use d to estimate H.)

5. Calculate the underlying time series Y_t for the $ARIMA$ model after removing the fractional factor d.

   (a) Use the infinite auto-regressive filter: $Y_t = (1-B)^d X_t = \sum_{s=0}^{\infty} b(s) B^s X_t = \sum_{s=0}^{\infty} b(s) X_{t-s}$, where $b(s) = \Pi_{k=1}^{s} \dfrac{k+d-1}{k} = \dfrac{\Gamma(s-d)}{\Gamma(-d)\Gamma(s+1)}$ and X_t is the observed data after stationary adjustments.

   (b) Choose a large s to stop the summation. (In our case we choose $s = 150$).

6. Adopting traditional time series methodology, find the most adequate $ARIMA$ model for the de-fractionalised data Y_t and estimate the parameters.

   (a) Apply the Box-Jenkins Method.

   (b) Provide a set of time series models from $ARMA(0,0)$ to $ARMA(6,6)$ and estimate all models in parallel.

   (c) Save the selected polynomials for the $ARMA$ ($\Phi(B)Y_t = \Theta(B)\epsilon_t$) process as $\widehat{\Phi}(B)$ and $\widehat{\Theta}(B)$.

7. Using the $ARMA$ filter, calculate $\widehat{Y}_t = \dfrac{\widehat{\Phi}(B)}{\widehat{\Theta}(B)} X_t$ to find the pure $ARFIMA(0, d, 0)$ process.

8. Re-estimate fractional parameter d from the "pure" process $\widehat{Y}_t$. Repeat steps 3 to 8.

9. Stop the algorithm until d converges.

3.2.4 Convergence and stability

To test the convergence and stability of our algorithm, we conducted another two experiments on the simulated data in Table A.4.5 and Table A.4.3. In Table A.4.5, we tested the three different Hurst exponent methods together with their average on simulated data to monitor the convergence of our algorithm in controlled environments. We simulated 10000 data for three types of $ARFIMA$ models: with AR term, with MA term and with AR and MA terms. The 'Steps' in the second column of the table indicate the current loop of the recurrent algorithm as we proposed above. One can see all simulated data convergence in terms of d (thus, the convergence of the Hurst exponent) with all methods. Convergences happen quickly and are usually stable within the first 10 steps. The right-most column provides the average error for each method at each step, which, again, indicates that the AV and OT method work better in most of the cases. Note, if we take the Hurst exponent as the average of the three methods, we find decent results in terms of overall error and convergence speed.

In Table A.4.3, we provide the empirical results for our algorithm on estimating the parameters for the $ARFIMA$ model applied to simulated data. These are the fitted results from six individual random runs of our algorithm. Based on a simulated $ARFIMA$ process with positive and negative Hurst exponents, as well as different ranges of the fractional parameter, the simulated time series all have 10000 time steps, which is similar to the real data we have. One can find that the random selection of our fitting procedure provides adequate performance with acceptable average errors from an overall perspective.

3.2.5 Estimation intervals and stability

In order to measure the stability of the estimation and the significance of the estimated parameters[1], we propose the following testing algorithms following a similar approach done in [55] for $ARIMA$ models. By applying a direct filter with the estimated parameters, we re-sampled the identical but independent distributed residuals to regenerate a new sequence with similar memory properties. Then, the fractional parameter d is estimated again from this newly generated process. Such a procedure repeats several times, and one can obtain an empirical distribution from the procedure for d. More specifically, the algorithm for finding confidence intervals and t-Statistics is as follows:

1. Find the optimal model $ARMA$ ($\Phi(B)Y_t = \Theta(B)\epsilon_t$) model and the corresponding Hurst exponent using the algorithm we explained earlier.
2. Construct the residual term as: $\epsilon_t^d = (1-B)^d \left(\frac{\Phi(B)}{\Theta(B)}\right)(X_t - \mu)$ where X_t is the original input data and ϵ_t^d is the residual with estimated fractional parameter d.
3. Regenerate independent and identical normally distributed random number ϵ_t with the same variance as ϵ_t^d.
4. Put ϵ_t back to the equation in step 2, and generate a new sequences $\dot{X}_t$ using the same $\Phi(B), \Theta(B), \mu, d$ and σ as previous estimated.
5. Re-estimate $\dot{d}$ from $\dot{X}_t$ using the previous algorithm.
6. Repeat step 3-5; one will have simulated candidates of $\dot{d}$'s for d.
7. If $\dot{d}$ follows a stationary distribution, one can calculate the $5\% - 95\%$ estimation intervals by taking the 5 and 95 percentiles from $\dot{d}$'s.

In the next section we apply the estimation method we developed in the current section on commodity prices and report the results.

[1]Note, the statistical significance for $ARMA$ parameters is reported in TableA.4.10 and A.4.11 as AICs and BICs.

3.3 Application on commodity price and market

3.3.1 Commodity price data sets and their stationary

In this section, we fit commodity price data into the ARFIMA models and select the best model for each product using the algorithm above. The data we chose are the daily trading data from Bloomberg for the past 20 or 30 years. For all individual commodities, we used nearby (front month) futures price data, except for the commodity index which is a weighted index (GSCI). We used the commodity index or future prices ranging from industry, oil, metal, livestock and non-livestock. Unlike the stock prices, the prices for commodities and the relevant financial derivatives are less liquid and may not be completely hedge-able. This may be due to the properties of physical goods, storability and convenient yield.

Table A.4.6 summarises the statistical property of the selected data together with the p-values from the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests and the augmented Dickey–Fuller (ADF) tests for stationary processes. One can tell that most of the data are heavily skewed with fat tail when compared to normal distribution whose skewness is 0 and kurtosis is 3.

On the other hand, from the p-values, it is not hard to see in general that all of the prices are not stationary. This creates some trouble in our work as our algorithm assumption that requires the data to be stationary. Thus, instead of studying the price data directly, we chose to work with the difference of the prices to indicate the volatility of the data. The log differences (log return, defined as $Y_t = \log X_t - \log X_{t-1}$) and the first difference (increments, defined as $Y_t = X_t - X_{t-1}$) are typical choices for empirical study as they fit the underlying assumption for most of the financial theories and are easy to work with. Also, it is mentioned in [79] that the absolute values of the log return are a good measure for data's long memory. In the following part, we will only work with absolute values of log returns and the incremental data. Table A.4.7 shows the KPSS and ADF tests for the processed data (Diff: increments; Log: log difference), except for 'CME FEEDER CATTLE INDEX' and perhaps 'BRENT CRUDE OIL INDEX' in KPSS tests; all other commodities exhibit stationary property after processing.

It is obvious that the infinite filter in step 5 from the algorithm above converges to a small enough number that it can be ignored. As a matter of fact, in our practice,

we found $s = 150$ to be an adequate choice if not too large (See Figure 3.3.1 for a comparison on $b(s)$'s convergence with difference d values). However, in Step 6, we reckoned that it is not feasible to choose the best ARIMA model for the time series data manually even if we followed the Box-Jenkins method. This is because, essentially, we are dealing with a data set with 40 different types of commodities, and it is too time-consuming so that is not practically comparable with several models. To solve this problem, after individually studying the data samples, we recognised usually a model among $ARMA(p, q), p, q = 1, ..., 6$ as good enough in terms of modelling, simulation and prediction. Hence, for the filtering procedures, for each commodity, the algorithm runs separately on each of these possible combinations for a total of $7 \times 7 = 49$ times. All fitting results and statistical testing parameters were saved in local files. After that, all models were estimated, and a third program compared the Akaike or Bayesian information criteria (AIC and BIC). Statistically speaking, the only uncertain part of our estimation, given the fixed data set, is from the $ARMA$ model estimation; thus, AIC and BIC are the proper statistics by which to measure the goodness of estimation. From each realisation of the estimation, we chose the model according to the AIC and BIC under the condition that convergence in d occurred. One consideration of this implementation is the calculation efficiency and running time. However, we believe this approach does not necessarily extend the running time for the whole algorithm, as each model filter is estimated separately and does not require any information from the previous steps. So, one can implement them in parallel and saving a significant amount of computation time.

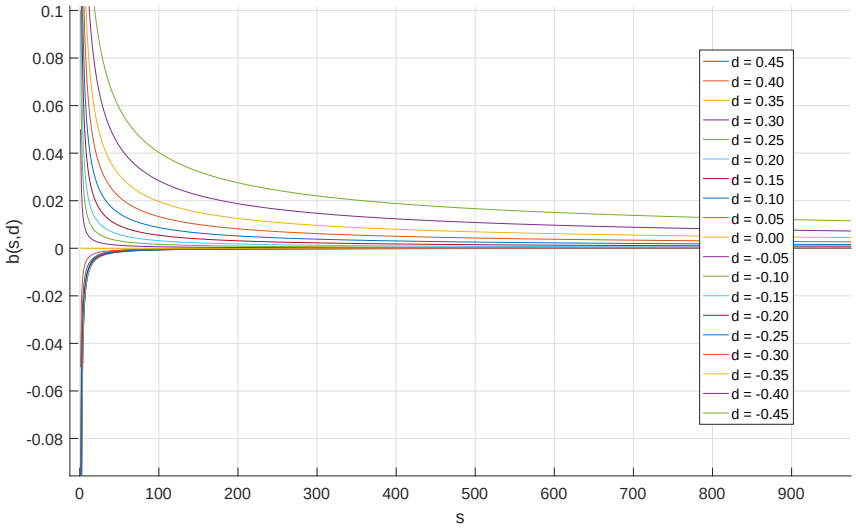


Figure 3.3.1: The convergence of $b(s)$ for $d = -0.45$ to $d = 0.45$

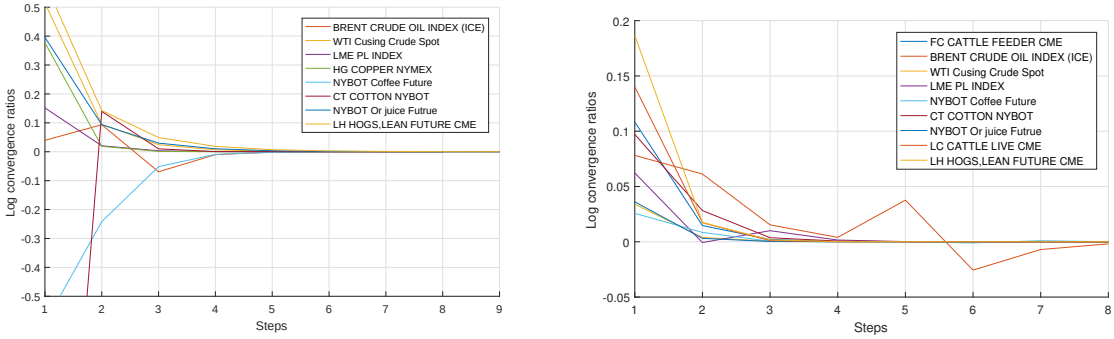


Figure 3.3.2: Log plots of convergence ratios (Left: increments Right: Log difference)

3.3.2 Hurst exponent and ARFIMA model results of commodity price

In Table A.4.1 and Table A.4.2, we demonstrate the Hurst exponent results from our algorithms. The optimal $ARFIMA$ models we found using our algorithm are also provided in Figure 3.3.7. The Akaike and Bayesian information criteria are provided in the last column as local statistic indexes, and the estimation intervals are also provided[2]. We finally chose the best $ARFIMA$ models based on the smallest

[2]Note, following the algorithm in subsection 3.2.5, we re-sampled 1000 times for each selected optimal model in order to find the corresponding intervals. A case study is made on increment of the 'BRENT CRUDE OIL INDEX' data, where we demonstrate the histogram of 10000 re-sampling

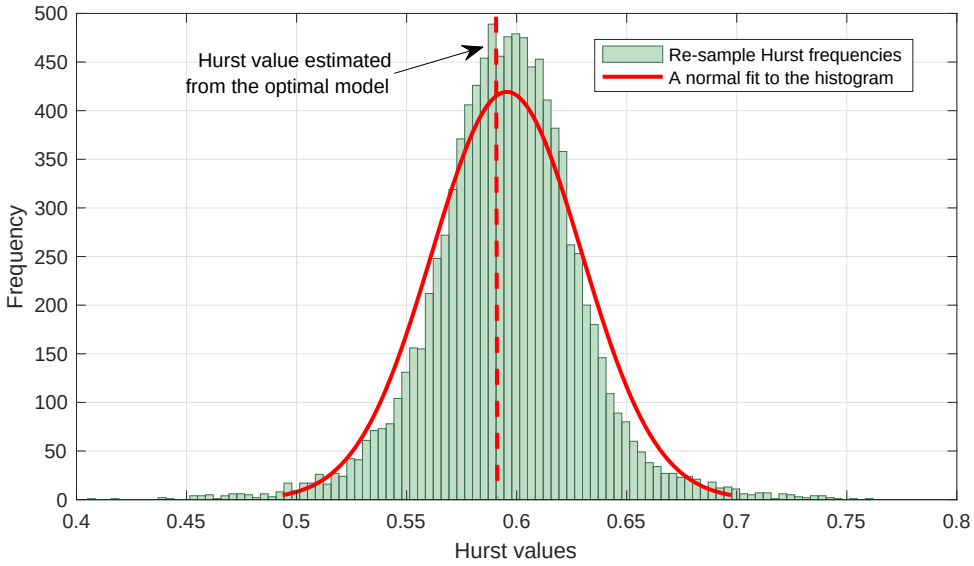


Figure 3.3.3: Histogram of re-estimated 10000 Hurst exponents for the increment of 'BRENT CRUDE OIL INDEX' price

AIC/BIC from all of our selections. Note, a vertical comparison between AIC/BICs from different commodities is meaningless and should be avoided, as these are local statistical indicators calculated from specific models with the corresponding commodity price data. Within a commodity, a smaller AIC/BIC means the underlying model is more efficient in terms of the optimal log-likelihood function values and model complexities.

Since we use the absolute log return as suggested in [79], it is more an index for measuring the volatility of commodity price movements rather than to forecasting it. Since the process of the absolute values can only provide information about the price movements, and not its directions, it can be only used to detect the process's long memory. In other words, the best $ARFIMA$ model we found based on this measure can only be used to simulate and forecast the volatility index itself but not the price of the commodity. However, this is good enough if our focus is only on the long memory side and may be useful to explain the repeating patterns in the absolute log return auto-correlation for most of the commodities; see Figure 3.3.4

Hurst values in Figure 3.3.3. As one can tell, the re-sampling shows a bell-shaped curve, which fits normal distribution fairly well, and we observed the mean values of the re-estimations as close to the initial estimations we made for the Hurst exponent using our algorithm. We have made the same observations for all other commodities. However, given that the values limitation for Hurst is $[0, 1]$, for Hurst values close to 0 and 1, the bell shape is cut at the boundary.

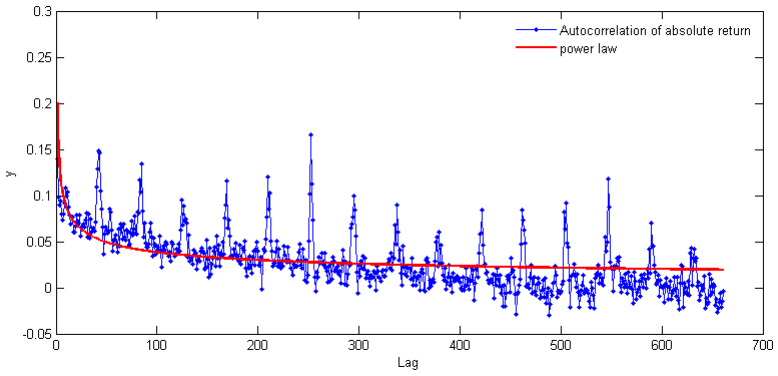


Figure 3.3.4: Long-memory in the auto-correlation for commodity price data (Lumber data here)

for example. From Table A.4.1, we find all the commodity variability exhibits LRD, and they are persistence as all Hurst exponents are larger than 0.5. This result is close to the results in [79], where a different method is applied on slightly different processed data. Among the results, one can see the CME Feeder Cattle Index, CBOT Corn Future, CBOT Oats Future and WCE Canola Future have the smaller Hurst exponents than others while CBOT Wheat Future has the largest.

As the previous comparisons focus more on the data memory, for $ARFIMA$ model estimation, we used the same approach again with the increments of the price. With the first difference price data, our models economically make sense and can be used to simulate and forecast future scenarios. In Table A.4.2, we sort the Hurst indexes in ascending order, and find that, despite the Feeder Cattle price and LB Lumber CME, all other livestock prices have a lower Hurst exponent when compared to the crops. A similar observation of CEV models is made in [5]. In [5], the author derived the constant elastic volatility (CEV) model by using the Ito theorem for a group of similar agriculture data. Then the model parameters are estimated using the maximum likelihood method. In [5], the value of the CEV model parameter differentiates the live- and non-livestock commodities due to the market demand elasticity. Note, the same linear trend with slightly lower values can be found if the estimation of Hurst component is made based on average values of the re-sampling samples. Such a method is mentioned in subsection 3.2.5 when estimation interval is discussed. This trend, again, indicates the stability of our estimation as well as the statistical significance of the linear trend.

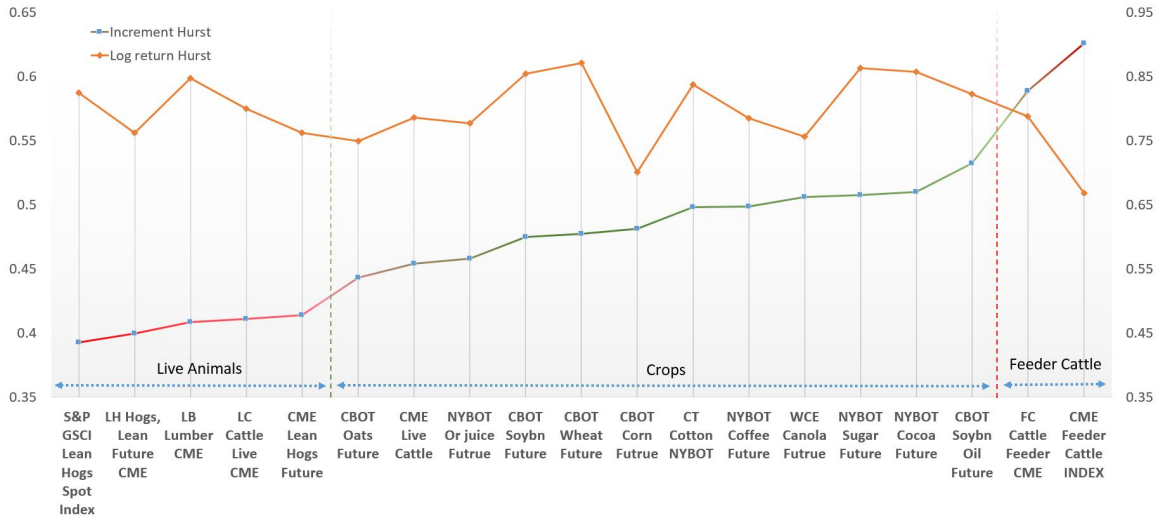


Figure 3.3.5: Hurst exponents for commodity price data (log return vs increments)

A one-to-one comparison can be found in Figure 3.3.5 for the results from different data process methods, where we observe a linear increment for the first difference results from livestock to crops but random jumps for the log return results if sorted in the same order. One can see, in the increments data, the Lean Hogs Spot Index has the lowest Hurst exponent while, both Feeder Cattle Index and Future have the largest ones. Thirteen out of the 19 commodities have a Hurst index smaller than 0.5, which indicates they exhibit anti-persistence or mean reversion property; i.e. a growth in data is more likely to be followed by a decay. Note, a $H < 0.5$ does not indicate short-range dependence; see Definition12 for details. CT Cotton NYBOT and NYBOT Coffee Future have Hurst exponents that are quite close to 0.5, which suggests market efficiency and short-range dependence in the increment data themselves.

For commodities with Hurst exponents larger than 0.5, e.g. NYBOT Sugar Future, NYBOT Cocoa Future, CBOT Soybean Oil, Future FC, Cattle Feeder CME and CME Feeder Cattle INDEX, we see another distinctive difference between crops and Feeder Cattle. Since the underlying data are the increments of price, the LRD actually indicates the long-lasting and positive effects of autocorrelations in the changes of the price; more specifically, an increase in price is more likely followed by a price increase.

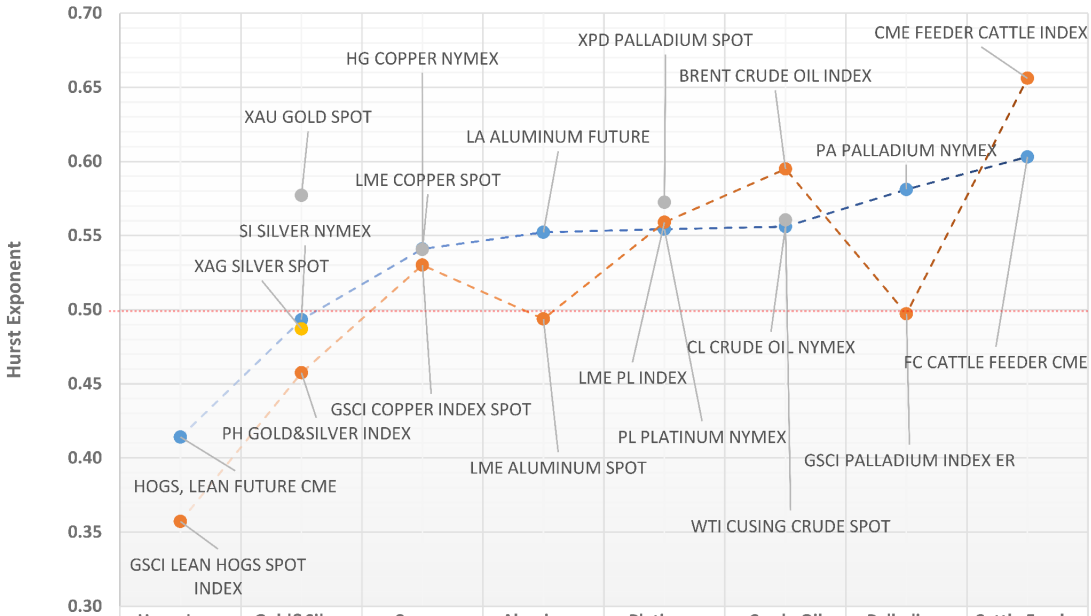
3.3.3 Hurst exponent and ARFIMA model results of commodity market

Table A.4.9 is the output example for 'CBOT Corn Future' increment data. In total, 49 models are estimated using the algorithm above. The 'NoS' stands for Number of Steps required for d to converge to a stable value. A minimum boundary of 9 is set for stability concerns. One can see that all models converge rapidly and different models do provide significantly different estimations of the fractional parameter. Eventually, an $ARFIMA(5, -0.0218, 5)$ is chosen by sorting the 'AIC/BIC' in an ascending order, and the Hurst exponent is estimated as the average of the best 3 models. We also provide Figure 3.3.2 for some of the convergence test. In these figures, the logarithm of the convergence ratios calculated from both increments data and log difference are presented. We select some of the most representative commodities, and one can see they all converge very quickly, usually within 9 steps.

Table A.4.10 and A.4.11 provide the final results. In these tables, we list the best $ARFIMA$ model for each of the 40 commodities and report their model parameters in the incremental form and log difference form. We see the algorithm works with all commodities, and they usually converge very quickly (a minimum step of 9 is used in all recursions for stability reasons). The details of the estimation of each commodity is saved in the local files (provided by request) with the convergence test as well as statistical test values for each parameter. In Figure 3.3.7, we give an overview of the Hurst exponent for all commodities and the chosen optimal model for each of them. As one can see in the vertical bar plot, the orange bars indicate the Hurst exponent for the absolute values log difference data and the blue bars are for the first difference data. We find the log return generates larger Hurst exponents than the increments for all data; these results are consistent with the results from [107, 79], where the Hurst exponents for agriculture commodities are calculated in another two difference methods. Furthermore, there is no obvious pattern in the optimal $ARFIMA$ models our algorithm picked, but we see the consideration for a total number of 49 models is indeed needed, as some of the commodities require rather large models to cooperate their dynamics.

For those commodities which have index, future or spot prices, we compare their Hurst exponents in figure 3.3.6. Despite the fact that a future contact and an index

for the same commodity may not necessarily written on the exact same product, we can still observe a trend where future prices tend to have a closer values to 0.5 when compared with the index prices, and spot prices always have a larger Hurst exponent than the future and index prices. Recall that a Hurst exponent close to 0.5 means more liquidity in the contract based on the assumptions from financial mathematics. It is not difficult to understand future prices as derivatives prices are traded more frequently than indexes; thus, they behave more randomly. Since the spot price of a product is mostly related to the product itself and all the seasonal facts affecting it, the larger values in Hurst exponents, which indicate high persistence, for the spot data also seem reasonable to us.



	Hogs, Lean	Gold&Silver	Copper	Aluminum	Platinum	Crude Oil	Palladium	Cattle Feeder
Future	0.4141	0.4932	0.5409	0.5522	0.5542	0.5561	0.5812	0.6031
Index	0.3571	0.4574	0.5300	0.4937	0.5590	0.5948	0.4972	0.6560
Spot		0.5770	0.5406		0.5724	0.5605		
Spot		0.4870						

Figure 3.3.6: Hurst exponents comparison among future and index

Another interesting discovery for commodity categorisation is that agriculture products tend to have smaller Hurst exponents than industry products if we sort the commodity with respect to their Hurst exponent calculated from increments. This pattern does not necessary hold when we examine the log difference data or if the Hurst exponent is calculated by the R/S method without the recursion. See

Table A.4.8, where we summarise the estimation results for the Hurst exponent with increments data and log difference data for both our method and the R/S method in on run; this pattern only shows in 'Hurst(Diff)'. In Figure 3.3.7, we color-coded all commodities with respect to their categories. From bottom to top, red is for livestock or animal, green is for non-livestock or crops, grey is for metal and orange is for industrial oil. On average, despite feeder cattle, there is a pattern where the Hurst exponent increase going from livestock to non-livestock to metal and eventually to industrial oil. For most of the cases, live- and non-livestock commodities have a Hurst exponent smaller than 0.5 for their first difference while metals and oils have Hurst exponents larger than 0.5. We believe this difference between commodities could be caused partly by their demand elasticity, as similar observations have been made in [5] for CEV model parameters on agriculture product data.

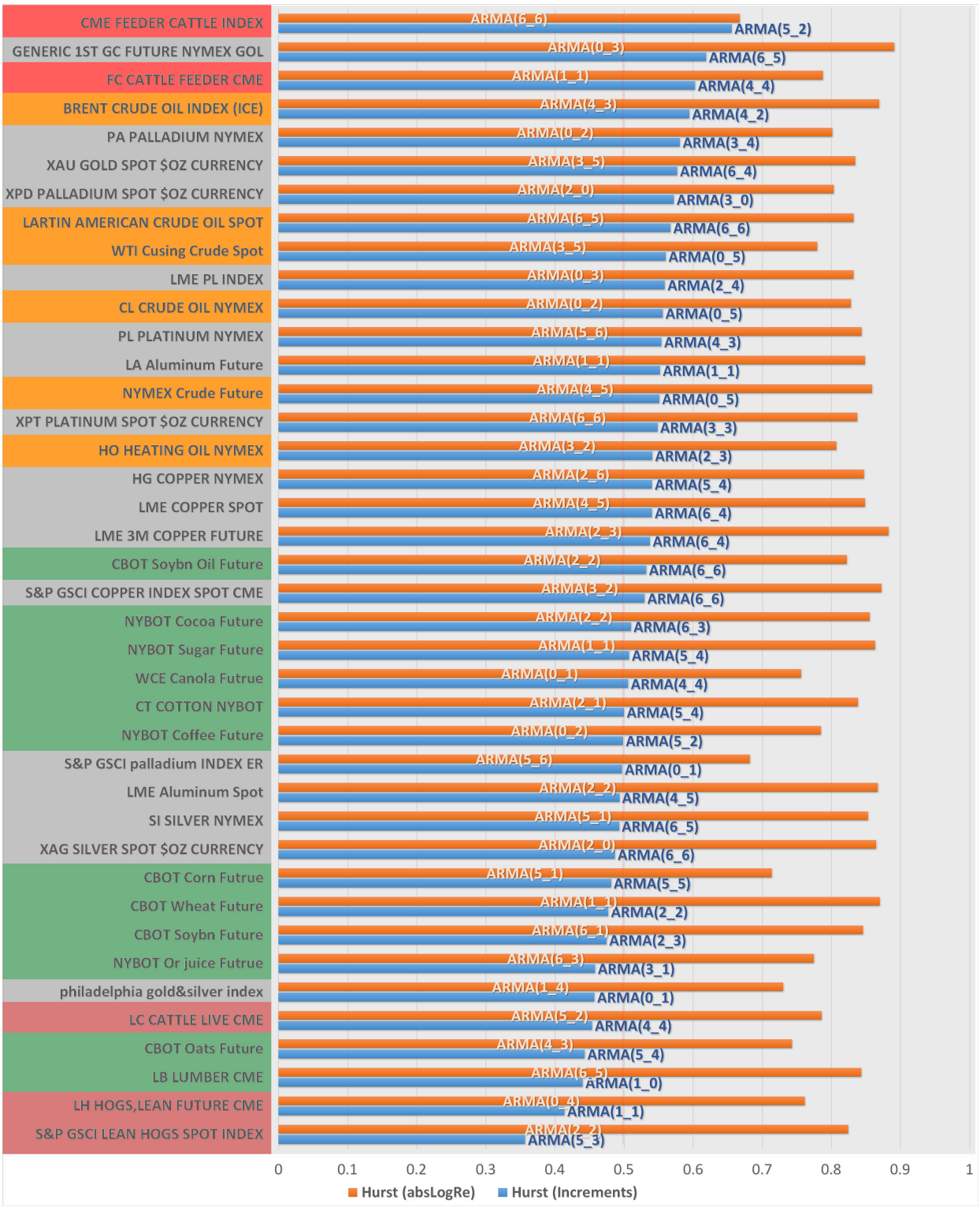


Figure 3.3.7: Fitting model results and Hurst exponents for all data

3.4 Interim Conclusion

The importance of data modelling is addressed in this chapter for answering Question 4.

Here, a new and stable algorithm to calculate the Hurst exponent was applied to commodity prices and markets. The unsuitability of the maximum likelihood estimator is demonstrated with likelihood function plots. Two Hurst exponent estimation methods were reviewed and a new one was proposed. A recursive method was then proposed by employing the relation between the Hurst exponent and the fractional parameter in ARFIMA models. Also, the stability of our proposed method was tested with artificial data for error control and with real data for statistical significance examination. The statistical significance calculation algorithm is proposed with bootstrap and ARMA filters.

Via our estimation method, we estimated the Hurst exponent of the real data. It was argued that the absolute log return of all studied agriculture products exhibited long-range dependence, while most of the increment prices are anti-persistent. These are true after a careful examination of statistical significance. Note that the estimation of model parameters follows a recursive algorithm that filters the inadequate models via the information criteria, namely AIC and BIC. Thus, statistical significance is already guaranteed within the selected candidate models.

Further observation was consistent with [107, 79]. In the first difference case (increments), we found results similar to the CEV models in [5], where one can distinguish between live-animal prices and crops' prices. For commodity markets, we also observed a trend where future prices tend to have a closer value to 0.5 and spot prices always have a larger Hurst exponent than the future and index prices; moreover, the increments of agriculture product tend to have smaller Hurst exponents than industry products.

In addition, the optimal ARFIMA models for each product are also estimated by comparing the information criteria. As one can see in Figure 3.3.7, one can have good combinations of $ARFIMA$ models for AR and MA steps varying from 0 to 0.7. The diversification of model selections also indicates the efficiency of the algorithm: models with larger AR and MA terms that may not be necessarily better than smaller models.

Chapter 4

Feasibility analysis for a new type of agriculture insurance

In this chapter, we propose a framework for designing a new product in the agricultural insurance market. Then we test our findings on the UK agricultural index prices. Two problems are considered here, based on two different sides of an agricultural index insurance market. First, we consider the insureds (demand side) consisting of farmers, farm production retailers and wholesalers. We set an optimal problem where the risk of the insurer's global position is minimised. Second, we consider the insurers (supply side) who have introduced this product, consisting of insurance companies and investors investing in this new product. We set up a portfolio management problem to show the profitability of investing in this market.

Besides the theoretical results of this chapter that show how one can find the optimal insurance contracts on farm index prices, we have found that such a missing market on the UK farm index prices is a feasible practice. Since agricultural insurances are major tools for farm risk management due to the undeveloped derivative markets in the UK, the results of this chapter can be used by insurance companies to design a new market for farm index insurance products. While the issue of agricultural insurance pricing and design (or perhaps derivatives) is of a great concern to policymakers, investors and insurance companies, an insurance market for agricultural index prices in the UK is non-existent. To the best of our knowledge, this is the first work addressing this matter in a scientific study.

Managing volatile prices is one of the farming industry's biggest problems. Inelastic demand and supply means a small increase in supply can lead to a larger fall

in prices. However, this is a fundamental fact that is often overlooked by academics and policymakers. Increasing technology use requires increased investment, and investment requires confidence in the future. As volatility increases, farmer confidence drops. Farmers need more predictable income to invest in new technology, and bankers need reliable income to support a growing business.

To break this roadblock, farmers need a simple, affordable and low-risk tool to help them manage the effects of volatile prices. This means they need a risk management tool designed and built for farmers, rather than for financiers. The outcome of this work is the introduction of a product to cover the risk of volatile prices among the UK farm businesses. This product helps a wide range of farming businesses protect themselves from volatile costs and risk of fluctuations. It is all based on public indexes from government organisations such as the AHDB[1] and DEFRA[2].

The reasons we write this Chapter include, first, there is no good market for agricultural insurances on price fluctuations in the UK. Second, the current arable derivatives in ICE (the Intercontinental for OTC markets, taken over by London International Financial Futures and Options Exchange in 2002) are not insurance, as insurances are not for speculation (must be issued at a certain amount for each client) and they have different markets. The third reason is Brexit and the removal of the common agricultural policy which contributed in direct payment to the UK farmers and could eventually damage livestock businesses in Wales.

In the following, we take care of the two ends of a new market: demand and supply sides. Demand for insurance is made by risk-averse farmers and farm-product consumers who are looking for lower risk, and the supply side are the investors and underwriters who invest in this business and issue insurances for making higher profits.

On one hand, customers should be convinced that the new insurance is a good enough product to manage the risk of volatile prices, and, on the other hand, investors should be convinced that there is a portfolio of the products that makes enough profit. In the following, we show a schematic picture of how both sides of the market can interact.

[1]Agriculture and Horticulture Development Board
[2]Department for Environment, Food and Rural Affairs

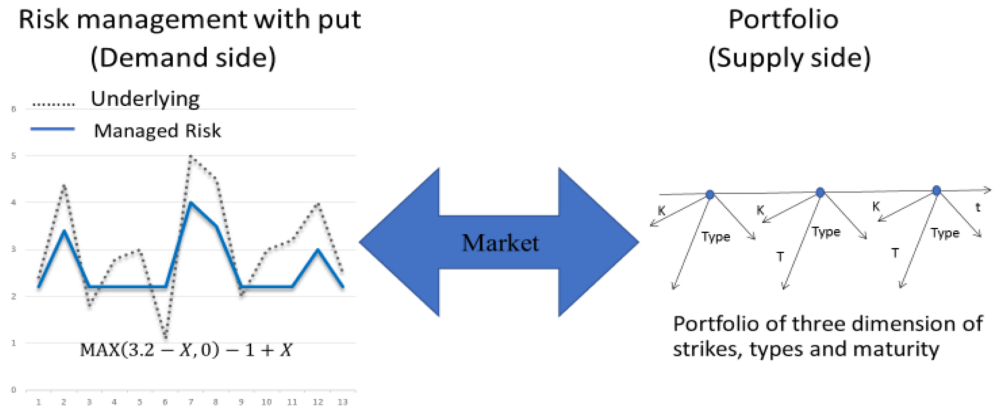


Figure 4.0.1: An illustration of risk transfer

The products that we consider are put or call spreads. For simplicity, the examples we supplied here are put bear spread or call bull spread with two European-style vanilla options: one at (or close to) the money and one out of the money. This is to mimic the pay-off structure of two-layer stop-loss insurances. The differences between our policy and a traditional insurance contract include:

1. The policies are only to protect against price falls (or rises). No yield or other factors will be taken into consideration for the final settlement. The settlement only happens once at the end of the contract.

2. The based index prices are publicly available data published by a mutually trusted third party.

3. Both the settlement price and the price of the policies (premium) can be adjusted monthly if not weekly. Quotes for different contracts can be generated automatically in real-time.

4. Both the settlement price and the price of the policies (premium) will be adjusted monthly once the new index price data are available.

5. The client cannot negotiate premiums, and all settlements are executed automatically.

6. The pricing method requires underwriters to set a preferred risk levels (KPIs like Loss Ratios, ROIs, Premium rate).

7. The pricing method is based on the traditional actuarial pricing methods, if no market or claims have been made before.

Following this setting, we are particularly interested in two subjects: the strike prices, i.e. the retention levels for such two-layer stop-loss policies, and the prices of the policies.

In the rest of the chapter, we introduce the data sets first in Section 4.1. The underlying models are introduced in Section 4.2 together with policy specifications. We discuss the optimal solutions to the demand side in Section 4.3 and provide sensitivity analysis for the theoretical results in Section 4.4. Section 4.5 is dedicated to the supply side discussions which leads to Section 4.6 and eventually Section 4.7 to report all the results. A short conclusion is made in the last section.

4.1 Data Set

The underlying assets are agricultural products ranging from Dairy, Livestock and Arable products. The monthly data set is from AHDB database. A summary of the data's characteristics is provided in Table 4.1.1. Note, linear interpolation is used for missing data points.

Product	Metric	Data range	Most recent Price	#Available points	Average return
Feed Wheat	£/tonne	Jan2005-Feb2017	144.20	146	0.82%
Feed Barley	£/tonne	Jan2005- Feb2017	121.70	146	0.65%
Milling Wheat	£/tonne	Jan2005- Feb2017	148.80	143	0.62%
OSR	£/tonne	Jan2005- Feb2017	362.80	146	0.79%
Pig	p/kg	Jan2005-Jan2017	147.96	145	0.31%
Potato AP	£/tonne	Jul2005- Feb2017	145.99	140	1.54%
Potato	£/tonne	Jul2005- Feb2017	250.19	140	3.38%
Milk	p/litre	Jan2005-Dec2017	26.21	144	0.29%
Deadweight Cattle	p/kg	Jan2006-Dec2017	352.90	132	0.48%
Lamb Deadweight SQQ	p/kg	Jan2006-Dec2017	386.90	132	0.53%

Table 4.1.1: Data characteristics summary

As one can see, for the average return (monthly return calculated average), all product prices are with up trends. This indicates the underlying market is a good market for issuing stop-loss policies to hedge against price drop events.

In Table 4.1.2, we also provide the co-variance matrix, where one can see, except for the potato, all other products are highly correlated. Indeed, the potato normally follows a very different behaviour; for instance, potato prices are very volatile. This can be due to high demand for the potato product and derivatives in the food market e.g., restaurants largely consume potatoes on daily basis. This can be also a sign of speculation in the potato market, although we cannot provide any evidence for this in this study.

4.2 Models and Contract

Let us consider the index process (I_t) is following a geometric Brownian Motion process

$$I_t = \mu I_t dt + \sigma I_t dW_t.$$

This process has an explicit form that can be expressed as follows:

$$I_t = \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right).$$

Here μ, and σ are constants for growth shifts and volatility of the asset. W_t is a standard Brownian motion. Let us consider a time horizon T at which we want to introduce a loss variable and make an insurance contract to hedge against the risk of the losses.

In order to be able to define the loss variable, we need to predict the future index value. Let us denote the value of the prediction by $\hat{I}_T$. This value is in terms of today's prices. There are different methods for prediction, including reduced-form auto-regressive models e.g., ARMA, ARIMA or ARFIMA models, or machine-learning algorithms e.g., neural network algorithm. Finding the best method for forecasting is out of scope of this research, so we just assume the predicted value $\hat{I}_T$ is based on all information available today, which is just a constant number. Based on the predicted value $\hat{I}_T$ one can introduce two loss variables: one from the farm

	Feed Wheat	Feed Barley	Milling Wheat	OSR	Potato	Potato AP	Milk	Deadweight Cattle	Lamb Deadweight	Pig
Feed Wheat	1.00	0.98	0.96	0.90	0.53	0.59	0.73	0.48	0.43	0.61
Feed Barley	0.98	1.00	0.93	0.87	0.52	0.56	0.68	0.42	0.36	0.51
Milling Wheat	0.96	0.93	1.00	0.84	0.46	0.52	0.77	0.49	0.40	0.64
OSR	0.90	0.87	0.84	1.00	0.47	0.57	0.73	0.62	0.62	0.69
Potato	0.53	0.52	0.46	0.47	1.00	0.95	0.21	0.27	0.15	0.24
Potato AP	0.59	0.56	0.52	0.57	0.95	1.00	0.34	0.42	0.33	0.40
Milk	0.73	0.68	0.77	0.73	0.21	0.34	1.00	0.75	0.55	0.86
Deadweight Cattle	0.48	0.42	0.49	0.62	0.27	0.42	0.75	1.00	0.76	0.81
Lamb Deadweight	0.43	0.36	0.40	0.62	0.15	0.33	0.55	0.76	1.00	0.70
Pig	0.61	0.51	0.64	0.69	0.24	0.40	0.86	0.81	0.70	1.00

Table 4.1.2: Correlations between products

product seller (like farmer) as $L = \left(\hat{I}_T - e^{-rT} I_T\right)_+$ and one from the farm product buyer $L = \left(e^{-rT} I_T - \hat{I}_T\right)_+$. Without loss of generality, in the theoretical setting, we consider the loss from the farm product seller's point of view. The loss distribution is given as follows (see appendix A.5 for details):

$$F_L(x) = \begin{cases} 0, & x < 0 \\ N\left(\frac{\left(\mu - r - \frac{1}{2}\sigma^2\right)T - \log\left(\frac{\hat{I}_T - x}{I_0}\right)}{\sigma\sqrt{T}}\right), & 0 \le x < \hat{I}_T \\ 1, & x \ge \hat{I}_T \end{cases}.$$

Here, N is the cumulative distribution function of a standard normal distribution. The graph of $F_L(x)$ is depicted in Figure 4.2.1.

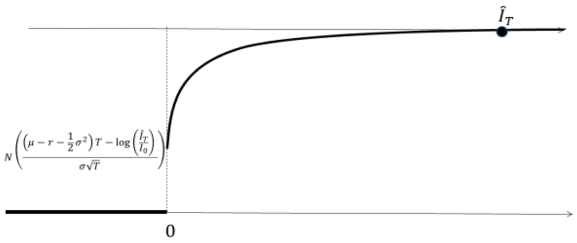


Figure 4.2.1: Loss densities

4.2.1 Premium

Using the risk-free approach to pricing, we can price any contract $H = h(I_T)$ by using the risk-free probability measure Q given by

$$Price = e^{-rT} E^Q\left(h\left(I_T\right)\right) = e^{-rT} E\left(\frac{dQ}{dP} h\left(I_T\right)\right),$$

where

$$\frac{dQ}{dP} = e^{\left(\frac{m^2}{2\sigma^2} - \frac{m}{2}\right)T} \left(\frac{e^{-rT} I_T}{I_0}\right)^{-\frac{m}{\sigma^2}},$$

and $m = \mu - r$. However, since the indexes are not traded in the market, we should replace $\mu - r$ by $\lambda\sigma$ where λ is the market price of risk, whereby we assume it is a constant number. From this, one can show that

$$\pi\left(L\right) = E\left(\frac{dQ}{dP} h\left(I_T\right)\right) = \int_0^1 \mathrm{VaR}_t\ \left(h\left(I_T\right)\right) d\Gamma\left(t\right),$$

where $\Gamma(t) = N\left(N^{-1}(t) - \frac{|m|\sqrt{T}}{\sigma}\right) = N\left(N^{-1}(t) - \lambda\sqrt{T}\right)$ and N is the cumulative distribution function of a standard normal distribution.

Note that, by incorporating the market price of risk, we have:

$$F_L(x) = \begin{cases} 0, & x < 0 \\ N\left(\left(\lambda - \frac{1}{2}\sigma\right)\sqrt{T} - \frac{\log\left(\frac{\hat{I}_T - x}{I_0}\right)}{\sigma\sqrt{T}}\right), & 0 \leq x < \hat{I}_T \\ 1, & x \geq \hat{I}_T \end{cases}$$

In designing an optimal insurance contract, we should consider the moral hazard risk and consider a contract that rules it out. The literature on actuarial mathematics deals with such a problem in the following manner by considering that both insurer and insured should feel losses.

We assume that a contract X should be so that both X and $L - X$ are increasing in L. Hence, the following assumption is necessary.

Assumption 1. *We consider contracts $X = k(L)$, where k belongs to the following set:*

$$C = \{k : R_+ \to R_+ \mid k(x) \text{ and } x - k(x) \text{ are non} - \text{decreasing in } x\}.$$

4.3 Demand side: Insured

In this part, we set up an optimal insurance problem and seek an optimal solution. For that, we assume that the insured is a risk-averse agent whose risk is measured by a distortion risk measure ρ on the set of non-negative random variables defined by:

Definition 16. *A distortion risk measure is a mapping defined on the set of all random variables that can be represented as follows:*

$$\rho(X) = \int_0^1 VaR_t(X)\, d\Pi(t)$$

Here $\Pi : [0,1] \to [0,1]$ is a non-decreasing function so that $\Pi(0) = 0$ and $\Pi(1) = 1$.

Remark 1. *This form can easily be shown to be equivalent to what we introduced earlier in 1.2 where* $\Pi(x) = 1 - g(1-x)$.

This family of risk measures includes very important examples, e.g., Value at Risk with $\Pi(t) = 1_{[\alpha,1]}$ or Conditional Value at Risk with $\Pi(t) = \frac{t-\alpha}{1-\alpha} 1_{[\alpha,1]}$.

The insurer's global loss is the part of the loss that is not covered by the insurance added to the amount that is paid for by the premium i.e.,

$$Global\ loss = L - X + \pi(X).$$

Since distortion risk measures are cash invariant, then the risk of the global loss is $\rho(L-X) + \pi(X)$. To study insurance premiums, we set an optimal insurance design problem as proposed in [3] and [4]:

$$\min_{K \in C} \rho(L - k(L)) + \delta\pi(k(L)),$$

for a risk-loading factor $\rho \geq 1$ that is used by an insurance company. Using the marginal indemnification function method (MIF) introduced in [3] and [117] (and further in [4]), this problem can be re-written as follows:

$$\min_{0 \leq k' \leq 1} \int_0^1 \left(\delta\left(1 - \Gamma\left(F_L(t)\right)\right) - \left(1 - \Pi\left(F_L(t)\right)\right) k'(t)\right) \mathrm{dt}.$$

Here, k' is the derivative of k. The optimal solution is then given by $X = k(L)$, where

$$k'(t) = \begin{cases} 1, & 1 - \Pi\left(F_L(t)\right) > \delta\left(1 - \Gamma\left(F_L(t)\right)\right) \\ 0, & 1 - \Pi\left(F_L(t)\right) \leq \delta\left(1 - \Gamma\left(F_L(t)\right)\right) \end{cases}$$

Assumption 2. *We assume that there are* $a, b \in (0,1)$ *so that* $1 - \Pi(x) > \delta(1 - \Gamma(x))$ *on* (a,b) *and* $1 - \Pi(x) < \delta(1 - \Gamma(x))$ *on* $(0,a) \cup (b,1)$.

The following proposition is clear enough and needs no proof (for illustration of what is happening, see Figure 4.3.1).

Proposition 4. *Assumption 2 holds for* ρ*=VaR and CVaR. For VaR,* $b = \alpha$ *and* a *is the solution to* $\delta\Gamma(t) = 1$*, which gives*

$$a = N\left(N^{-1}\left(\frac{1}{\delta}\right) + \lambda\sqrt{T}\right).$$

For CVaR, a is the same as in the case of VaR, and b is the solution to the following equation:

$$\frac{1-t}{1-\alpha} = \delta\Gamma(t) = \delta N\left(N^{-1}(t) - \lambda\sqrt{T}\right).$$

In Figure 4.3.1 we have shown how a, b can be found for VaR and CVaR.

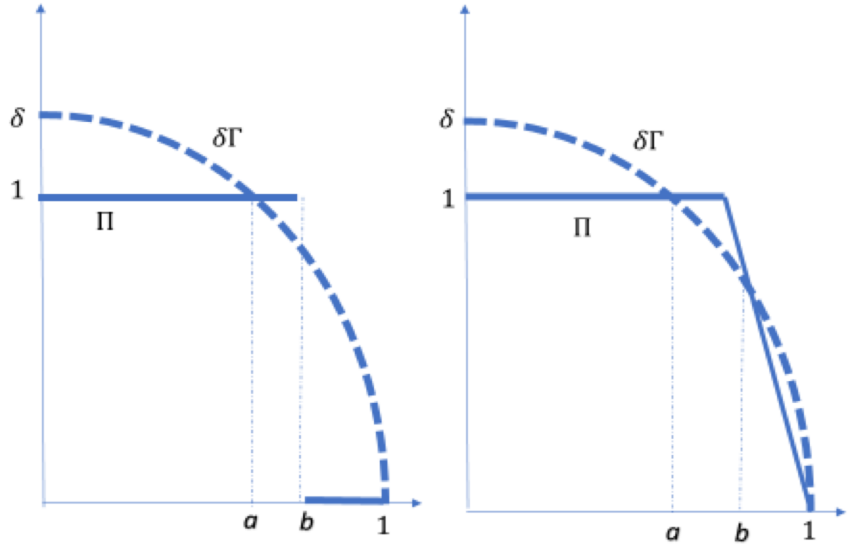


Figure 4.3.1: Possible solutions for cases with VaR and CVaR

Definition 17. *A two-layer policy with lower and upper retention levels l and u, respectively, is defined as*

$$f(x) = \begin{cases} 0, & x < l \\ x - l, & l \le x < u \\ u - l, & x \ge u \end{cases} \tag{4.3.1}$$

The following proposition is again clear, and, for illustration purposes, one only need to see Figures 4.2.1 and 4.3.2 (a similar detailed proof is provided in [3]).

Proposition 5. *If $F_L(0) < a$, the contract is a two-layer policy with lower retention and upper retention levels as*

$$I = I_0\left(1 - \exp\left(\sigma\left(\lambda - \frac{1}{2}\sigma\right)T - \sigma\sqrt{T}N^{-1}(a)\right)\right),$$

$$U = I_0 \left(1 - \exp\left(\sigma\left(\lambda - \frac{1}{2}\sigma\right)T - \sigma\sqrt{T}N^{-1}(b)\right)\right).$$

The existence of the optimal solution and its form depends on

$$F_L(0) = N\left(\left(\lambda - \frac{1}{2}\sigma\right)\sqrt{T} - \frac{\log\left(\frac{\hat{I}_T}{I_0}\right)}{\sigma\sqrt{T}}\right).$$

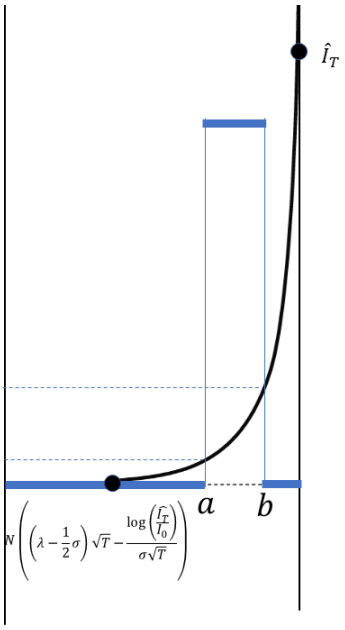


Figure 4.3.2: Illustrations for how the layers works

As one can see, there are lots of parameters at play when designing and pricing an optimal contract, including volatility, interest rate, market price of risk, aversion parameter, etc.

4.4 Sensitivity analysis

In order to better understand the impact of the parameters involved in designing and pricing the optimal insurance contracts we found in the previous section, we present some sensitivity analysis based on the model just provided.

These analyses are from syntactical data and are for demonstration purpose only. The parameters we used are

- Start index price: $I_0 = 100$.
- Estimated index price at time T: $\hat{I}_T = 100$.
- Underlying market volatility: $\sigma = 0.1,\ 0.12,\ 0.14,\ \ldots,\ 2$.
- VaR and CVaR criteria: $\alpha = 99\%$.
- Market price of risk factor: $\lambda = 1$.
- Risk loading factor: $\delta = 1.01,\ 1.02,\ \ldots,\ 1.99$.
- Risk-measures: VaR and CVaR.

Note that, in all formulas, we have found that, for the parameters (upper and lower retention levels), r is irrelevant once we know λ. Since in the sensitivity analysis, we fix λ, we really do not assume any value for r. However, if one needs to retrieve μ (which is not relevant here), one needs to know r and use $\lambda = \frac{\mu - r}{\sigma}$. First, we plot the upper and lower bounds for both VaR and CVaR risk measures with different volatilities. In Figure 4.4.1, one can see how the bounds are changing with respect to the growth of risks. There are three interesting observations. First, both upper bounds (u) and the lower bound (l) increase with respect to risks (i.e., σ). This is due to the anticipated price movements in the future, the larger the volatility, the more price deviations can be expected from current prices later. Thus, to reduce risk, the algorithm requires the boundaries of the extreme values in very rare events, e.g. $\sigma > 1$. Second, as we observed, the same trend of lower boundaries is observed for both measures. This is due to the value of a in the calculation of u in both scenarios. However, a slightly higher upper bound for the VaR compared to CVaR indicates that CVaR is capturing more risks than VaR - hence, the tight protection interval. Also, in the final point, we notice that, in our parameter ranges, the condition of $F_L(0) < a$ is always satisfied. The decreasing values for $F_L(0)$ proves that, when risks increase, more probabilities are allocated to larger risks from a cumulative distribution function point of view.

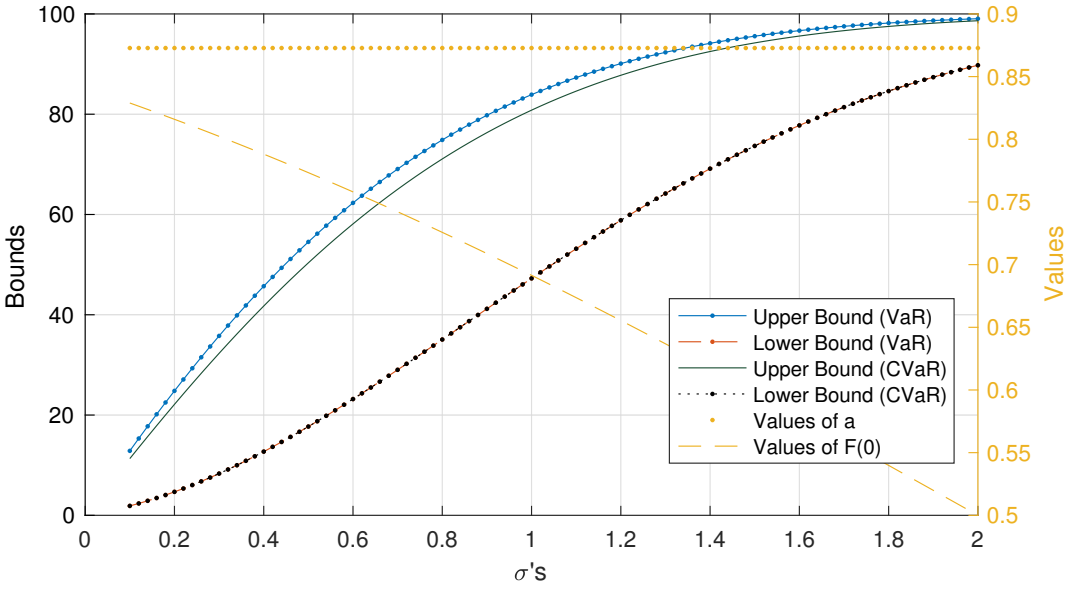


Figure 4.4.1: Sensitivity analysis for volatility with VaR and CVaR

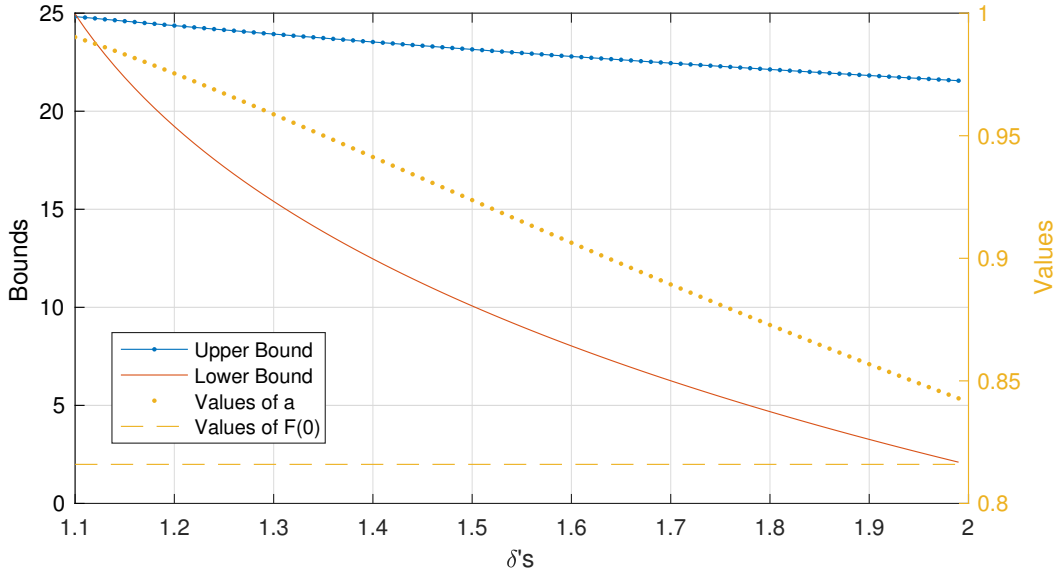


Figure 4.4.2: Sensitivity analysis for risk loading with CVaR

Now let us see if an underwriter has a different risk-loading requirement. We know, for the same risk appetite, larger risk loadings indicate undertaking more risks under the same risk-exposure setting. Figure 4.4.2 demonstrates this phenomenon by changing $\delta's$ from smaller to larger values. Note $F_L(0) < a$ still holds for all simulations. The low optimal bound decreases as δ increases, this means taking more

and more risks, so the probabilities of claims happening in this range are getting greater.

Lastly, we look at the contract length. Recall, our algorithms finds the optimal bounds by minimising global risks, and, for our policies' assumptions, the longer a policy lasts, the more risk exposure it endures. With $F_L(0) < a$ holds for all simulation ranges, we present the analysis for optimal bounds with respect to policy length in Figure 4.4.3. As expected, one can see the reducing of protection intervals as policy length increases from 1 year to 5 years. Interestingly, for longer contracts, the upper bound becomes lower than the low bound, which indicates the unavailability of the optimal solution due to unprecedented potential risk exposures in the long run.

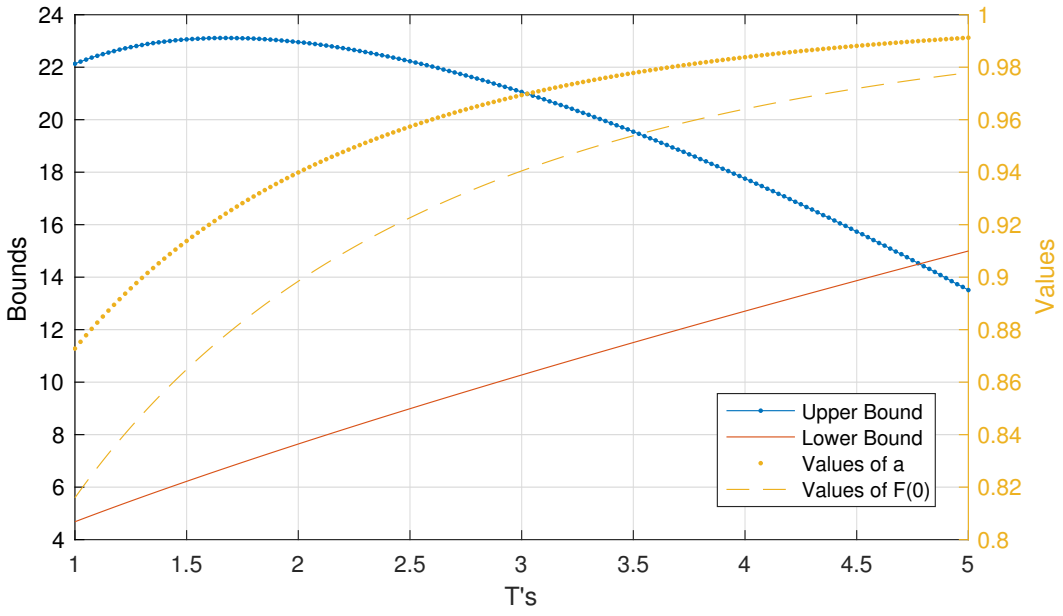


Figure 4.4.3: Sensitivity analysis for policy length with CVaR

These analyses summarise the behaviour of our algorithms from the demand side; it is now clear that the model behaves as anticipated, and the simulated results are consistent with the design purpose of the models.

4.5 Supply side: Investors

As we discussed in the previous part, the main insurance contracts we need to consider are spreads (which are eventually two-layer policies). We illustrate the payoff and

mechanism of put and call spreads in Figure 4.5.1.

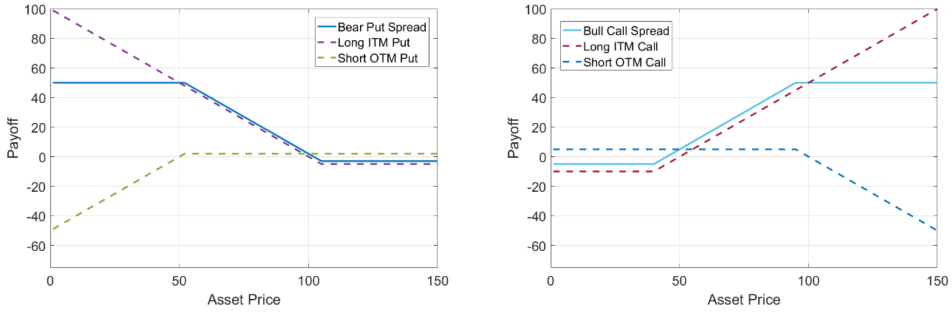


Figure 4.5.1: Spreads payoffs (left: bear put spread; right: bull call spread)

In Section 4.3, the retention levels l, u are functions of the index volatility σ, (I_0), market price of risk (λ), risk aversion parameter (α), forecasting methods $(\hat{I}_T)$ and the maturity (T). It is not clear, though, if the market participants have homogeneous assessments of these parameters. That is why finding an optimal contract can be impossible in this way. However, the optimality of two-layer policies now is clear to us; therefore, we consider a set of these products and base our analysis on finding the optimal retention levels by using a different set of criteria. To construct a diversified portfolio, we assume the underwritten contracts are different in contract length and settlement price. More specifically, for each underlying asset, we shall issue 10 maturities and 5 strike prices. Thus, our policies range from 3 months to 12 months with settlement prices at 75%, 80%, 85%, 90%, 95% of the current market asset prices. In total, we have 10 underlying assets, each with 30 variations. Thus, there are 300 different configurations for all policies. These spreads are all issued to hedge against price fluctuations.

Now, the problem is, first, how do we find the right strike prices (retention levels), second, how to price these contracts and, finally, how to form a profitable portfolio of them.

In order to reduce the underwriters' risks, we take one step further and apply forecasting algorithms to the product prices. More specifically, we use the Seasonal ARIMA Time Series Model to forecast asset prices at any given time.

Definition 18. A Season ARIMA model - $ARIMA(p,d,q) \times (P,D,Q)S$ - is given as

$$\Phi(B^S)\varphi(B)(I_t - \mu) = \Theta(B^S)\theta(B)w_t$$

where the non-seasonal components are the AR polynomials: $\varphi(B) = 1 - \varphi_1 B - ... - \varphi_p B^p$ and the MA polynomials: $\theta(B) = 1 + \theta_1 B + ... + \theta_q B^q$. The seasonal components are seasonal AR polynomials: $\Phi(BS) = 1 - \Phi_1 B^S - ... - \Phi_P B^{PS}$ and seasonal MA polynomials: $\Theta(BS) = 1 + \Theta_1 B^S + ... + \Theta_Q B^{QS}$. Where I_t is the underlying process, B is the lag operator, $w_t \sim N(0, 1)$ is the residual and $p,\ d,\ q,\ P,\ ,D,\ Q,\ S = 0, 1, 2, 3, ...$.

At each time step, we forecast the values of I_t for the same policy length and denote it as $\hat{I}_t$. Then $\hat{I}_t$ is adjusted by I_t to give the strike price base $S_t = min(I_t,\ \hat{I}_t)$. This is to avoid the forecast result yielding price rises as the purpose of the policy is to protect price falls (vice versa for bull spreads). If higher-than-current market prices are taken as strike price bases, the underwriters' risks could be amplified by the forecast algorithms rather than reduced.

The strike price base S_t is then applied to the percentages we just presented to yield the final upper bound price of the stop-loss policy. For simplicity, the lower bound layer ratio is chosen as 60 % of the base price S_t.

4.6 Policy pricing

However, due to incomplete market information and the instabilities involved in the model calibration to real data, one still needs a pricing model to determine the contract prices in the real world. In this section, we introduce a new pricing method that generates policies' prices based on historical data and the expectations from both the demand and the supply side. First, we state the fundamental rules behind our pricing algorithm:

First, premium is positively correlated with the amount of risks an underwriter takes; i.e. more risks indicate higher premiums.

This, translated in terms of contract length, means one can have,

Contracts that are closer to the current market price will create more risk for the insurer. For any two contracts, where the only difference is the maturity date, the longer one will be more expensive than the shorter one.

Also, in terms of strikes ratios, one can have,

The closer a contract is to the current market price, the more risk for the insurer. Thus, for any two contracts where the only difference is the strike, the one that is closer to the underlying asset price is more expensive than the other one.

It is straightforward to justify these statements under a fair and efficient market condition. It is obvious for the Black-Scholes framework that all statements stand. The Black-Scholes model provides the pricing formula for vanilla European-style options.

Note, our policy shares the same payoff as a spread. Equation 4.3.1 can be re-written as: $f(x) = \max(0,\ u - x) - \max(0,\ l - x)$, for $u \geq l$. Thus, under the Black-Scholes-Merton framework, the policy price is:

$$Price = N(-d_2)ue^{-rT} - N(-d_1)I_t - N(-d_4)le^{-rT} + N(-d_3)I_t \tag{4.6.1}$$

where $d_1 = \frac{1}{\sigma\sqrt{T}}[ln(\frac{I_t}{u}) + (r + \frac{\sigma^2}{2})T]$, $d_2 = d_1 - \sigma\sqrt{T}$, $d_3 = \frac{1}{\sigma\sqrt{T}}[ln(\frac{I_t}{l}) + (r + \frac{\sigma^2}{2})T]$ and $d_4 = d_3 - \sigma\sqrt{T}$.

4.6.1 Filtering mechanisms

So far, with the help of numerical methods, one can generate vast sets of potential option prices. However, we still do not know which one is optimal and has the correct (fair) price. In this section, we introduce a filtering method based on different perspectives for both sides of the market, so that we can arrive at a fair price that satisfies all market participants.

The market fair price of policies should balance between both the demand and the supply side. In measuring both sides' preferences, key performance indicators (KPI) play important roles. More specifically, to evaluate underwritten policies, insurers usually monitor loss ratio (LR) and return over investment (ROI). On the demand side, clients or policyholders usually are very sensitive to premium rate, in particular Spread Price vs Asset Price (SPvsAPs) . The Sharpe ratio (SR) is used by all sides of the market to determine the quality of portfolios. We define the calculation of KPIs as follows:

Loss ratio

$$LS = Claims/Premium$$

Return over investment

$$ROI = (Premium - Claims)/(u - l)$$

Premium rate

$$SPvsAP = Premium/Asset\ Price$$

Sharpe ratio

$$SR = (Expected\ return)/Volatility$$

While loss ratio indicates the frequency and size of the claims, return over investment measures the profitability of the polices. The premium rate indicates the fairness of policy prices and Sharpe ratios are usually used in a portfolio setting where a selection is needed among different underlying assets. Note here, $Claims$ is the amount of money an insurer pays to a client at the end of the contract and can be calculated as $Claims = \max(0, u - I_T) - \max(0,\ l - I_T)$. $Premium$ is the price of a contract that the insurer collects from a client, and its formula is provided by equation 4.6.1. $Asset\ Price$ is I_t, the price of underlying asset at the issuing of the contract. $Expected\ return$ is the expectation of ROI over all time periods. $Volatility$ is the standard deviation of $(Premium - Claims)$ over all time periods.

Thus, a filter algorithm is designed as follows to facilitate policy pricing

1. For each agriculture product, calculate spread prices following equation 4.6.1 and upper/lower retention bounds specifications with a wide range of implied volatilities (we use 150 equally incremental values from 0 to 35σ).

2. The spread price calculations are done for all available data sets historically, so that we can calculate the empirical performances of KPIs in the next step.

3. Calculate the KPIs for each configuration at each time step; this forms a large KPI pool for all potential policies that waiting to be selected as the ones with fair prices.

4. Underwriters and clients input preferred KPI values, such as setting a loss ratio smaller than 50%, choosing a return over investment larger than 5% per year, a premium rate smaller than 8% and the largest possible Sharpe ratio.

5. Filter out the un-qualified policies based on KPI values as in Figure 4.6.1.

6. Report the results, as to whether such a contract exist or not. For smaller simulation ranges or very tight filters, it is possible for one to end up with unsuitable configurations.

7. Rank the initial filtering results based on Sharpe ratios, and choose the final pricing configuration.

For example, if we keep μ as the estimated mean from the data and choose 150 equally incremental values from 0 to 35σ volatility hence 150 configurations, for the same policies (contract with the same underlying assets, up and low bounds and same maturity length), one will have 150 different prices at any given time in history following formula 4.6.1. With these generated prices, the calculation of KPIs is then applied to each of them. We list the KPIs for these 150 possible scenarios in Figure 4.6.1 with seven years of historical data and all KPIs are taken as seven years' average.

As expected, average loss ratios decrease as volatility goes up, expected ROIs (EROI) have a slight increment and premium rates have the same patterns as average contract prices. Note that, from top to bottom, the plots in Figure 4.6.1 where KPI filters are applied step by step result in smaller and smaller configuration ranges. Finally, in the bottom plot, we arrive at the initial filter results, a set of results that satisfies both the demand and supply side. These initial results are then again ranked by their Sharpe ratios to yield the best pricing configuration. Figure 4.6.2 summarises the process. Since the pricing mechanism relies on the KPI conditions, one cannot guarantee that the final policy prices can be found. However, in this research, we are issuing around 300 contracts, and only a few of them are unqualified.

To summarise, our pricing model inputs and outputs at each stage are listed in Table 4.6.1.

4.7 Empirical results for portfolios

One of this chapter's purposes is to show the feasibility of the existence of a profitable agriculture insurance market. A method is developed to test this theory based on the UK's commodity data as introduced in the previous section. The steps of the methodology is proposed as follows:

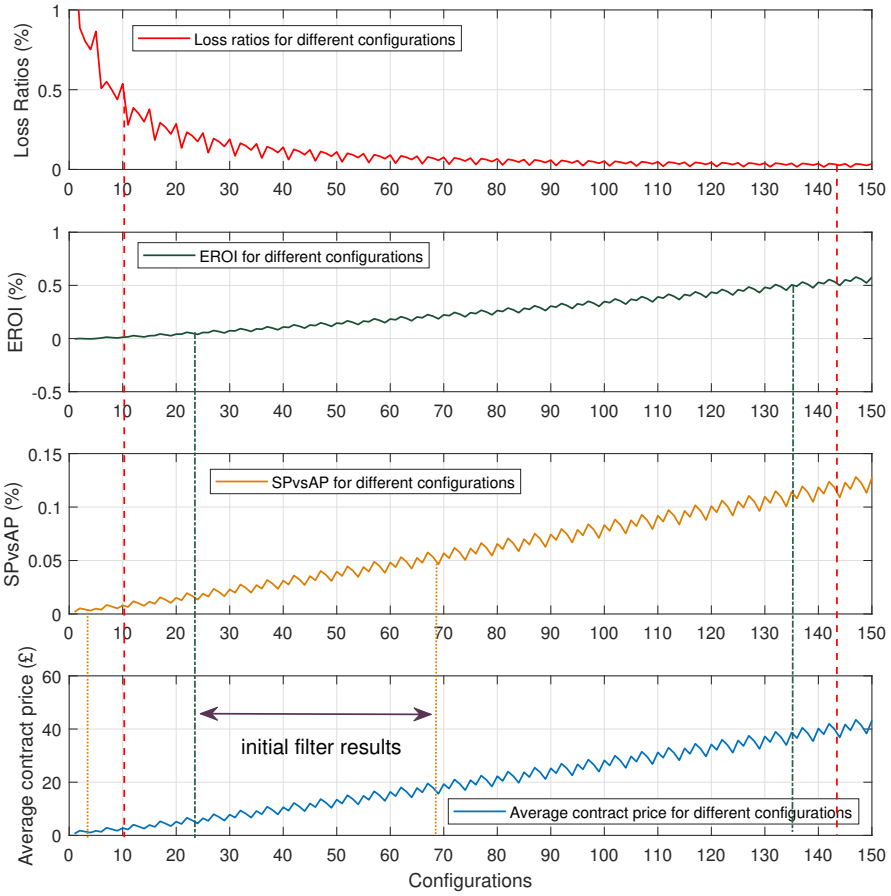


Figure 4.6.1: KPI filtering for one contract

Figure 4.6.2: The pricing process

Initial inputs	Internal estimations	Simulations	Outputs
Historical assets price	Implied volatility	Historical policy prices	Policies
Spread specification	Market price of risks	Random weights	Policy prices
Time to maturity	KPI filters	Simulation size	Efficient frontier
Risk-free rate	Policies		KPIs
Close derivative price	Potential policy prices		Capital allocations
Policy specification			
KPI ranges			

Table 4.6.1: Model inputs and outputs

1. We first set reasonable KPIs for the potential two-layer policies.

2. Then the pricing algorithm introduced previously is applied. For those contracts that can be priced, we calculated all their historical prices.

3. In the third step, the corresponding policy historical returns are calculated. These values are used in the portfolio return and volatility calculations.

4. Then, random combinations of the policies are generated to mimic possible capital allocation.

5. After finding the covariance matrix among all contracts, we calculated portfolio returns and risks, simulated the efficient frontier and found the greatest Sharpe ratio.

6. Repeat the previous two steps multiple times and stop the algorithm under one of the two conditions: if the number of total simulations passed a defined value or if the greatest Sharpe ratio stopped increasing for a relatively long simulation time. (In our practice, we chose to let the simulations stop after running 10^8 times)

7. If, from our simulations, the Sharpe ratio is greater than other financial products, we determine that such an insurance policy is finical more attractive to investors when compared to other financial products.

In following these steps and carrying out a simulation study next, the parameters we used for the final results are

- Product: ones listed in Table 1
- Date range: recent seven years of monthly data: October 2010 to October 2017
- Settlement prices: 75%, 80%, 85%, 90%, 95% of current asset market prices
- Stop prices: 60% of current asset market price
- Contract length: 3, 4, 5, 6, 8, 12-month
- Loss ratio: $\leq 60\%$

- Return over investment: $\geq 1\%$
- Premium rate: $\leq 8\%$
- Sharpe ratio: the larger the better
- Portfolio simulations: 10^8 sets of random weights for each maturity
- One-year risk-free rate: 0.05%
- Compare assets: FTSE recent 12-month performance
- Hedging direction: price falls or price rises

In principle, such configurations should provide us with 50 policies for each maturity at any given month in the previous seven years (Note, we ignore the contracts that have been issued most recently as we do not have future price data to close these policies; thus, their return data are not available). After running the algorithm, 10^8 random portfolios were generated with prices calculated through the KIP filters. We chose to summarise the simulation results into three scenarios for portfolio return and standard deviation (risks or volatility). For each maturity, we present portfolio performance by finding the largest Sharpe ratio, smallest portfolio standard deviations and greatest portfolio returns. These scenarios are recorded in Table 4.7.1.

Bear Spreads, hedge against price falls							
Scenarios	**Maturities**	**3**	**4**	**5**	**6**	**8**	**12**
Best Sharpe Ratio	**Avg. return**	0.64	0.44	0.33	0.25	0.17	0.07
	Std.	0.09	0.07	0.05	0.04	0.03	0.05
	Sharpe Ratio	**7.21**	**6.07**	**6.06**	**6.20**	**5.02**	**1.55**
Best Standard Deviation	**Avg. return**	0.61	0.42	0.31	0.24	0.16	0.07
	Std.	0.09	0.07	0.05	0.04	0.03	0.04
	Sharpe Ratio	**7.08**	**5.93**	**5.94**	**6.12**	**4.88**	**1.55**
Best Return	**Avg. return**	0.71	0.48	0.35	0.27	0.18	0.08
	Std.	0.12	0.10	0.07	0.06	0.05	0.06
	Sharpe Ratio	**5.84**	**4.60**	**4.89**	**4.39**	**3.47**	**1.37**
Bull Spreads, hedge against price rises							
Scenarios	**Maturities**	**3**	**4**	**5**	**6**	**8**	**12**
Best Sharpe Ratio	**Avg. return**	0.34	0.16	0.13	0.12	0.09	0.07
	Std.	0.66	0.35	0.30	0.25	0.17	0.11
	Sharpe Ratio	**0.51**	**0.47**	**0.44**	**0.49**	**0.55**	**0.64**
Best Standard Deviation	**Avg. return**	0.19	0.10	0.07	0.07	0.05	0.04
	Std.	0.39	0.23	0.19	0.17	0.11	0.08
	Sharpe Ratio	**0.48**	**0.43**	**0.38**	**0.40**	**0.47**	**0.52**
Best Return	**Avg. return**	0.46	0.26	0.15	0.13	0.11	0.08
	Std.	0.95	0.60	0.35	0.28	0.21	0.13
	Sharpe Ratio	**0.48**	**0.44**	**0.42**	**0.47**	**0.53**	**0.62**

Table 4.7.1: Sharpe ratios for policies

Table 4.7.1 shows the annualised[3] portfolio performance for different maturities. In our simulations, put bear spreads are used as two-layer insurance policies to hedge against price falls, and call bull spreads are used as similar policies to protect price rising risks. Recall Table 4.1.1, where we calculated the average monthly return of all ten products: all products exhibit prices' increasing trends during the data length we have obtained. This indicates, in terms of managing price fluctuation risks, that protecting price falls should be more profitable when compare to contracts protecting price rises. Nevertheless, the same conclusions are made here where the Sharpe ratios for bear spreads are significantly larger than bull spreads, particularly for shorter maturities from our simulation. Note, for the 12-month policies, both bear and bull spreads share the same returns in the 'Best Sharpe Ratio' scenario, and the differences in Sharpe ratios are only caused by the level of uncertainties.

[3]Annualised rate is calculated by the continuous compound method where Annualised rate = ((Current time during rate)‘$(1/n) - 1$), where (Current time during rate) * n = 1 year.

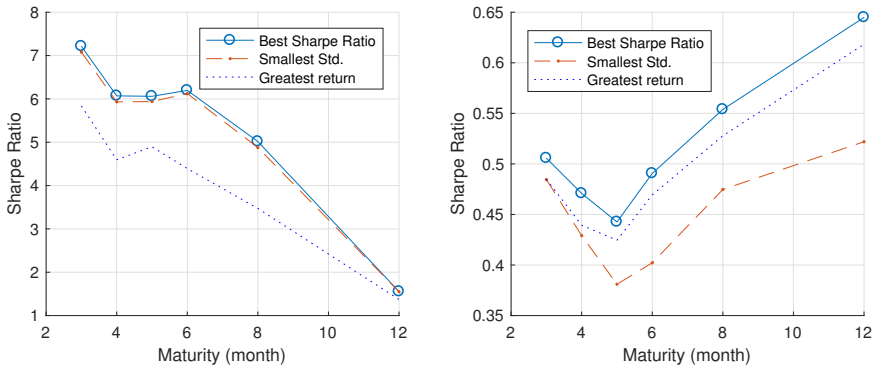


Figure 4.7.1: Sharpe ratios for different contract length. Left: bear spreads; Right: bull spreads

Figure 4.7.1 plots only the Sharpe ratios from Table 4.7.1. Interestingly, one can observe a decreasing trend in price fall protection contracts and an increasing trend in price rise protection contracts. Coincidentally, for all three scenarios, bear spreads converge to one point for longer contracts, but bull spreads start from almost the same point. The explanation behind these phenomena are beyond the scope of this research, but we think this is due to the lack of liquidity for longer contracts, which, together with the general up-going trend for prices, accumulate prices' uncertainties in long terms. However, these two plots indeed indicate risk capitals' potential preferences in both markets for price drop protections, insurers may issue more short-term policies than long-term policies due to larger Sharpe ratios and vice versa for price rise protections.

Lastly, we compared our portfolio results to financial market indices. The common indices for commodities are the DJ-UBS Commodity Index[4], Continuous Commodity Index[5] and S&P GSCI[6]. However, none of these indices are UK-specific, and all of them have smaller Sharpe ratios when compares to our portfolios. To our knowledge, the supply side of our portfolios are most likely to be insurance companies and risk capitals who are seeking long-term stable incomes. We think FTSE and its

[4]The DJ-UBSCI is composed of commodity futures contracts on physical commodities, traded on U.S exchanges. The only exception is aluminium, nickel and zinc which are traded in London (LME). This index is based upon relative trading activity of individual commodities.

[5]The CCI-index measure prices movements of 22 commodities. It has been live since 1957 and was earlier called the CRB Index.

[6]S&P and Goldman Sachs Commodity Index can be seen as a benchmark for investment performance in the commodity markets. S&P GSCI represents un-leveraged, long-only investments in commodity futures that is broadly diversified across the spectrum of commodities.

Index	12M Return %	12M Volatility %	Sharpe Ratio
FTSE 100	11.2	9.2	1.2
FTSE 250	14.3	8.2	1.7
FTSE SmallCap	18.0	5.7	3.1
FTSE All-Share	11.9	8.4	1.4
FTSE Fledgling	27.8	4.5	6.1

Table 4.7.2: FTSE market performances

sub-portfolios may be more suitable for this study. Table 4.7.2 lists the 12-month performances for some of the FTSE markets. As one can see, in our 12-month-maturity best case, a Sharpe ratio of 1.55 can be reached with a two-layer price fall protection that has an average return of 7% per year and a yearly volatility at 5%. It out-performs FTSE 100 and FTSE All-Share marker in terms of Sharpe ratios due to smaller volatilities. Meanwhile, for shorter-term contracts, most of the bear spreads generate better results than the FTSE market by providing risk capitals with greater annualised returns and less risky assets. Based on these observations, we think the price-based insurances, particularly the ones to hedge price falls, are good investment options and substitutions for insurers and risk capital.

4.8 Interim Conclusion

In this chapter, we present a new market potential in the UK's agriculture insurance sector by studying both the demand and the supply side. This contributes to answering question 3 and question 5.

On the demand side, a theoretical risk management framework is firstly proposed which proves a two-layer policy is the optimal setting to reduce overall risks. The method itself is based on [4] and [5], different from other approaches in the literature and aiming to find the optimal solutions for global risks. This both reduces the premiums a contract holder needs to pay and focuses on the risks that are not taken by underwriters. We first model the underlying index with a GBM process; then, the loss distribution is given by comparing the current index level to discounted future price. On the other hand, a risk premium is given following the risk-free approach. Combined with the marginal indemnification function method, we propose that a two-layer policy is the optimal solution to this problem on the demand side. The

conditions to be met and the analytic solutions for the upper and lower bounds are provided in 5.

On the supply side, our experiences with underwriters and risk capitals also suggest a two-layer policy is a preferable type of risk management tool. We proposed our solutions to two problems in this part: First, how to find the fair prices of these two-layer policies and, second, whether these contracts are worth investing for underwriter and risk capitals. For the first question, we proposed a pricing algorithm based on the KPI filters - i.e. suppliers' risk preferences - to find the acceptable prices for the two-layer stop-loss contracts. The pricing algorithm starts from deseasonalisation and a Black-Scholes framework. By simulating large numbers of possible scenarios, we obtained a pool of price candidates. These candidates are then filtered by KPI preferences to generate the final prices. From the Modern Portfolio Theory, we know portfolios with greater Sharpe ratios usually attract more investors. For the second question, real market index data are applied to the pricing model together with randomly generated market weights to simulate possible market behaviours for a portfolio or the contracts we want to underwrite. Observations from our results show that some of the simulated portfolios can over-perform in the FTSE market in terms of Sharpe ratios, average returns and portfolio risks. This indicates that, by using a two-layer protection policy, one reduces the overall risks from both the demand side and the supply side. We think such products are attractive to risk capitals and underwriters as insurance contracts as they are good investment opportunities and possibly good substitutions for major market indices.

Chapter 5

Conclusion

This thesis is about risk management in agricultural markets. More specifically, we discussed pricing frost insurance using quantitative analysis methods, parameter estimations for $ARFIMA$ model and pricing insurance contracts. All proposed methods and algorithms are tested with real data. The purposes of the thesis are to answer the questions posed in the first Chapter

1. How to measure the financial risks (particular the downside ones) for some risk exposure?

2. If a financial downside risk is anticipated, what are the tools one can use to control losses?

3. If a certain financial risk is identified, how can one price the protection cost for such a risk?

4. Can financial risks being forecast if one can simulate events that have long memories?

5. Can financial risks be reduced by studying the supply and demand of the underlying markets?

The first and second questions are answered in Chapter 1, where the fundamentals of risks, insurance and data models are introduced with real cattle price data. The simplification from general definition of risks to the negative financial risks is explained. The transformation from a general measurement of a histogram to a simple value risk measure is demonstrated. Definitions of insurance and reinsurance, as well

as different types of insurances are presented. Also, for the completeness of the thesis, we provide a simplified pricing method for insurance contract based on data from Example 1. Lastly, the continuous time stochastic model as well as discrete time series models are introduced. Chapter 1 lays down the foundation of the thesis and provides all the necessary background in mathematics and insurance.

In Chapter 2, we propose to price frost insurances with financial quantitative methods. We designed and priced a type of insurances strategy to protect the farmer from frost damage. In this work, we built the real loss variable from an indirect data set - temperature - with the help of distortion risk measures. This is due to the unavailability of the direct loss data and Moral hazard. We concluded that such an insurance strategies can be widely adapted to frost insurance where temperature data sets are available and a stop-loss style insurance is optimal in this framework as well.

In Chapter 4, we proposed a parameter estimation method for $ARFIMA$ processes. The method solves the problem in modelling long memory data and provide a more accurate Hurst exponent estimation. The method is applied to a set of commodity data wherein their best ARFIMA fittings are found. We discovered that, for most of the commodities, there exists long-memory in their index prices and a distinct trend is observed from products with different demand elastics.

In Chapter 4, we build a new risk management framework from both the demand and supply sides of an underlying market. In commodity risk management, the demand side of the market usually consists of producers and wholesalers, such as, farmers, food factories and supermarkets. They need risk management tools to protect against adverse events like price fluctuations. On the other hand, the underwriters and risk capitals are usually the supply side of these risk protection tools. A spread-style hedging strategies for the underwriters is discussed and a modified pricing method based on [5], [4] is proposed. We found that, based on 10 UK agriculture-product data sets, two-layer insurance polices are feasible and good alternatives for investment based on Sharpe Ratio analysis.

5.1 Future development

For further developments, we see the following potential problems from our current research. First of all, in Chapter 2, a better modelling from temperature and frost loss

can be developed with real experiments and more supporting data sets. The data we used in the thesis are from different sources, thus lacking consistency in both location and time. The modelling for losses can be significantly improved if more data sets are available and more complicated relations are introduced.

Also, in Chapter 2, we only studied a specific area of grove with no real frost damage recording data nor insurance claims data for comparison. This means we do not know if our pricing will be acceptable in reality, and the underwriters will also have no empirical data by which to test our models. Again, a real experiment in a controlled environment will improve the model as well as the underwriting of the policies.

In terms of Chapter 3, the biggest concern for us is the lack of explanations and implications for the two observations we made in the research. The first observation is the distinct differences for the Hurst exponent between live animals and crops. The values jumped when changing the category, possibly due to the differences in demands or trading patterns. The second observation is that spot prices always have a larger Hurst exponent than future and index prices. These discoveries are very useful when designing insurance contracts or trading on the market. This may be due to the fact that future markets are more liquid. Essentially, from the Hurst exponent, one can tell if a certain product's price is more likely to be repeated in long-term or short-term historical performance. As a matter of fact, the Hurst exponent is an important trading indicator in commodity trading.

The recursive algorithm we developed in Chapter 3 also suffers from a lack of efficiency. The ARMA filter is reconstructed each time in the loop, and the best model comes after testing 49 different models. This setting on its own limits the capacity of the model, as higher-order models may be better in specific scenarios, but this algorithm does not have it in the selection process. In the meanwhile, the algorithm does not carry any pre-model selection, thus becomes very computationally expensive with large data sets. All of these can be improved with a pre-estimation that reduces the selection pool of the optimal model recognition process.

For the third topic in Chapter 4, some clear improvements for our approaches include less complicated simulation models and more detailed study of the observations made from the simulations results. In the second half of Chapter 4, when we price the

two-layer policies from the supply side, the underlying pricing model is chosen as a Black-Scholes style. However, the pricing mechanism works with any type of pricing rule; with enough computational power, one can even price contracts by putting random numbers into the KPI filters. With this said, advanced models that can capture the real fair market prices more accurately will greatly accelerate the pricing process and provide one with more realistic prices. On the other hand, we observed some very interesting results from our final simulation. For instance, the shapes and patterns of the Sharpe ratios in Figure 4.7.1 have left us with even more curiosity in finding the hidden mathematical and economical reasons behind them. In the meanwhile, we think similar large-scale research will be worthwhile for more agriculture indices globally and perhaps other commodity indices beyond agriculture goods.

Bibliography

[1] Kenneth J Arrow. The Role of Securities in the Optimal Allocation of Risk-Bearing. *Review of Economic Studies*, 31(2):91–96, 1964.

[2] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[3] Hirbod Assa. Reinsurance optimal design with distortion risk measures and risk premiums. *Insurance: Mathematics and Economics*, 61:1–20, 2015.

[4] Hirbod Assa. Risk management under a prudential policy. *Decisions in Economics and Finance*, 38(2):217–230, 2015.

[5] Hirbod Assa. Financial engineering in pricing agricultural derivatives based on demand and volatility. *Agricultural Finance Review*, 76(1), 2016.

[6] Hirbod Assa and Keivan Mallahi Karai. Hedging, Pareto Optimality, and Good Deals. *Journal of Optimization Theory and Applications*, 157(3):900–917, 2013.

[7] Hirbod Assa and Meng Wang. A ruin based distortion risk measure and its applications. *Submitted*, 2015.

[8] Hirbod Assa, Meng Wang, and Athanasios A Pantelous. Modeling frost losses: Application to pricing frost insurance. *North American Actuarial Journal*, 22(1):137–159, 2018.

[9] Louis Bachellier. Theory of speculation. *The random character of stock market prices*, 1900.

[10] Jan Beran. A test of location for data with slowly decaying serial correlations. *Biometrika*, 76(2):261–269, 1989.

[11] Jan Beran. *Statistics for long-memory processes*. Routledge, 2017.

[12] Jan Beran, Yuanh Feng, Ua, Sucharita Ghosh, and Rafal Kulik. *Long-Memory Processes*. Springer, Berlin, Heidelberg, Berlin, 2013.

[13] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654, 1973.

[14] Karl Henrik Borch, Agnar Sandmo, and Knut K Aase. *Economics of insurance*, volume 29. Elsevier, 2014.

[15] Mohamed Boutahar, Velayoudom Marimoutou, and Leila Nouira. Estimation methods of the long memory parameter: Monte Carlo analysis and application. *Journal of Applied Statistics*, 34(3):261–301, 2007.

[16] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[17] Matthias Buchholz and Oliver Musshoff. The role of weaher derivatives and porfolio effects in agricultural water management. *Agricultural Water Management*, 146(C):1–6, 2013.

[18] Jun Cai and Ken Seng Tan. Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measures. *ASTIN Bulletin*, 37(1):93–112, 2007.

[19] Marc Chesney and Louis Scott. Pricing european currency options: A comparison of the modified black-scholes model and a random variance model. *Journal of Financial and Quantitative Analysis*, 24(03):267–284, 1989.

[20] Ka Chun Cheung. Optimal Reinsurance Revisited. *ASTIN Bulletin*, 40(1):221–239, 2010.

[21] Yichun Chi and Ken Seng Tan. Optimal Reinsurance under VaR and CVaR risk measures: a simplified approach. *ASTIN Bulletin*, 41(2):487–509, 2011.

[22] Yichun Chi and Ken Seng Tanb. Optimal reinsurance with general premium principles. *Insurance: Mathematics and Economics*, 52(2):180–189, 2013.

[23] Rama Cont. Long range dependence in financial markets. In *Fractals in Engineering*, pages 159–179. Springer, 2005.

[24] Nuno Crato and Philip Rothman. Fractional integration analysis of long-run behavior for US macroeconomic time series. *Economics Letters*, 45(3):287–291, 1994.

[25] M. Crouhy, D. Galai, and R.M. Mark, R. andf= Mark. *Risk Management*. McGraw-Hill Education, 2001.

[26] Angelos Dassios. *Ruin Probabilities*, volume 97. World scientific, 2002.

[27] Griselda Deelstra, Guillaume Plantin, et al. *Risk theory and reinsurance*. Springer, 2014.

[28] Allard E. Dembe and Leslie I. Boden. Moral Hazard: A Question of Morality? *NEW SOLUTIONS: A Journal of Environmental and Occupational Health Policy*, 10(3):257–279, 2000.

[29] Xiaohui Deng, Barry J. Barnett, Gerrit Hoogenboom, Yingzhuo Yu, and Axel Garcia y Garcia. Alternative Crop Insurance Indexes. *Journal of Agricultural and Applied Economics*, 40(1):223–237, 2008.

[30] Xiaohui Deng, Barry J. Barnett, Dmitry V. Vedenov, and Joe W. West. Hedging dairy production losses using weather-based index insurance. *Agricultural Economics*, 36(2):271–280, 2007.

[31] Peter George Muir Dickson. *The Sun Insurance Office, 1710-1960: The history of two and a half centuries of British insurance*. Oxford University Press, 1960.

[32] Kevin Dowd. *Beyond Value at Risk: The New Science of Risk Management, Frontiers in Finance Series*. John Wiley & Sons, 1998.

[33] Kevin Dowd. *Measuring Market Risk*. The Wiley Finance Series. Wiley, 2005.

[34] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935.

[35] Hans Föllmer and Alexander Schied. *Stochastic finance: an introduction in discrete time*. Walter de Gruyter, 2011.

[36] Lesław Gajek and Dariusz Zagrodny. Optimal reinsurance under general risk measures. *Insurance: Mathematics and Economics*, 34(2):227–240, 2004.

[37] Hélyette Geman. *Agricultural finance: from crops to land, water and infrastructure*. John Wiley & Sons, 2014.

[38] John Geweke and Susan Porter-Hudak. The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis*, 4(4):221–238, 1983.

[39] Clive W J Granger. The typical spectral shape of an economic variable. *Econometrica: Journal of the Econometric Society*, pages 150–161, 1966.

[40] Clive W J Granger. Long memory relationships and the aggregation of dynamic models. *Journal of econometrics*, 14(2):227–238, 1980.

[41] Clive W J Granger and Roselyne Joyeux. An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1(1):15–29, 1980.

[42] CT Haan, BJ Barfield, and JF Gerber. Risk analysis in environmental modifications. *Modification of the aerial environment of crops. Am. Soc. Agric. Eng. Monogr*, 2:30–51, 1979.

[43] James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384, 1989.

[44] Peter BR Hazell, Carlos Pomareda, and Alberto Valdés. *Crop insurance for agricultural development: Issues and experience*. IICA Biblioteca Venezuela, 1986.

[45] Tomoyuki Higuchi. Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, 31(2):277–283, 1988.

[46] John K Horowitz and Erik Lichtenberg. Insurance, Moral Hazard, and Chemical Use in Agriculture Insurance, Moral Hazard, and Chemical Use in Agriculture. *American Journal of Agricultural Economics*, 75(4):926–935, 1993.

[47] Jonathan R M Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.

[48] Darrell L Hueth and William H Furtan. *Economics of agricultural crop insurance: theory and evidence*, volume 4. Springer Science & Business Media, 2012.

[49] John C Hull. *Options, futures, and other derivatives*. Pearson Education India, 2006.

[50] Harold Edwin Hurst. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116:770–808, 1951.

[51] Ramiro Iturrioz, G Birbaumer, J Barrera Violeth, C Cardona Ayala, Cayón Salinas, F Boucher, A Espinoza Ortega, Roble Pensado Leglise, A Álvarez Macias, C Miranda, and Others. Agricultural insurance. Technical report, The World Bank, 2009.

[52] S Jaffee. Agriculture and rural development discussion paper. *World Bank, Washington DC*, 2005.

[53] Andreas Noack Jensen and Morten Ørregaard Nielsen. A fast fractional difference algorithm. *Journal of Time Series Analysis*, 35(5):428–436, 2014.

[54] Alan Kirman. Ants, rationality, and recruitment. *The Quarterly Journal of Economics*, 1(108):137–156, 1993.

[55] Soumendra Nath Lahiri. *Resampling methods for dependent data*. Springer Science & Business Media, New York, 2013.

[56] Beni Lauterbach and Paul Schultz. Pricing warrants: An empirical study of the black-scholes model and its alternatives. *The Journal of Finance*, 45(4):1181–1209, 1990.

[57] Blake LeBaron. Evolution and time horizons in an agent-based stock market. *Macroeconomic Dynamics*, 5(02):225–254, 2001.

[58] Antoine Leblois and Philippe Quirion. Weather-index drought insurance: an ex ante evaluation for millet growers in Niger. *Environmental & Resource Economics*, 57(27):527–551, 2011.

[59] Ivan M. Lee. Temperature insurance: an alternative to frost insurance in citrus. *Journal of Farm Economics*, 35(1):15–28, 1953.

[60] Ming Liu. Modeling long memory in stock market volatility. *Journal of Econometrics*, 99(1):139–171, 2000.

[61] A W. Lo. Fat Tails, Long Memory, and the Stock Market Since the 1960s, Economic Notes, 26 (2), 213-46. *International Library of Critical Writings In Economics*, 146(2):221–254, 2002.

[62] Andrew W Lo. Long-term memory in stock market prices. *Econometrica*, 59(5):1279–1314, 1991.

[63] Francis A Longstaff and Eduardo S Schwartz. Interest rate volatility and the term structure: A two-factor general equilibrium model. *The Journal of Finance*, 47(4):1259–1282, 1992.

[64] James D MacBeth and Larry J Merville. An empirical examination of the black-scholes call option pricing model. *The Journal of Finance*, 34(5):1173–1186, 1979.

[65] Olivier Mahul. The design of an optimal area yield crop insurance contract. *The Geneva Papers on Risk and Insurance Theory*, 24(2):159–171, 1999.

[66] Benoit B Mandelbrot. Une classe de processus stochastiques homothétíques à soi; application à la loi climatologique de H. E. Hurst. *Comptes rendus de l'Académie des Sciences*, 260(12):3274, 1965.

[67] Benoit B Mandelbrot. *The fractal geometry of nature*, volume 173. WH freeman, New York, 1983.

[68] Benoit B Mandelbrot and Murad S Taqqu. Robust R/S analysis of long run serial correlation. *Bulletin of the International Statistical Institute*, 48(2):69–104, 1979.

[69] Benoit B Mandelbrot and John W Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.

[70] Benoit B Mandelbrot and James R Wallis. Noah, Joseph, and operational hydrology. *Water resources research*, 4(5):909–918, 1968.

[71] Harry Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.

[72] Richard McCleary, Richard A Hay, Erroll E Meidinger, and David McDowall. *Applied time series analysis for the social sciences*. Sage Publications Beverly Hills, CA, 1980.

[73] Mario J Miranda. Area-yield crop insurance reconsidered. *American Journal of Agricultural Economics*, 73(2):233–242, 1991.

[74] Mario J Miranda and Joseph W Glauber. Systemic risk, reinsurance, and the failure of crop insurance markets. *American Journal of Agricultural Economics*, 79(1):206–215, 1997.

[75] Alberto Montanari, Renzo Rosso, and Murad S. Taqqu. Fractionally differenced ARIMA models applied to hydrologic time series: Identification, estimation, and simulation. *Water Resources Research*, 33(5):1035, 1997.

[76] Michael Norton, Eric Holthaus, Malgosia Madajewicz, and Nicole Peterson. Investigating Demand for Weather Index Insurance : Experimental Evidence from Ethiopia. Technical Report 845, 2011.

[77] Lysa Porth. *A Portfolio Optimization Model Combining Pooling and Group Buying of Reinsurance Under an Asset Liability Management Approach*. PhD thesis, University of Manitoba, 2011.

[78] Lysa Porth, Ken Seng Tan, and Chengguo Weng. Optimal reinsurance analysis from a crop insurer ' s perspective. *Agricultural Finance Review*, 73(2):310–328, 2013.

[79] Gabriel J Power and Calum G Turvey. Long-range dependence in the volatility of commodity futures prices: Wavelet-based evidence. *Physica A: Statistical Mechanics and its Applications*, 389(1):79–90, 2010.

[80] Roman Racine. *Estimating the Hurst Exponent*. PhD thesis, ETH Zurich, 2011.

[81] William Rea, Les Oxley, Marco Reale, and Jennifer Brown. Estimators for Long Range Dependence: An Empirical Study. *Fractals An Interdisciplinary Journal On The Complex Geometry Of Nature*, 3(1980):0–16, 2009.

[82] Valderio Reisen, Bovas Abraham, and Silvia Lopes. Estimation of Parameters in Arfima Processes: a Simulation Study. *Communications in Statistics: Simulation and Computation*, 30(4):787–803, 2001.

[83] Valderio A Reisen and Fabio A Fajardo. Robust estimation in time series with long and short memory properties. *arXiv preprint arXiv:1112.6308*, 39:17, 2011.

[84] Frans de Roon Rob W. J. van den Goorbergh and Bas J. M. Werker. Economic Hedging Portfolios. Technical report, 2003.

[85] Michael J Roberts, Nigel Key, and Erik O'Donoghue. Estimating the extent of moral hazard in crop insurance using administrative data. *Review of Agricultural Economics*, 28(3):381–390, 2006.

[86] Peter M Robinson. Semiparametric analysis of long-memory time series. *The Annals of Statistics*, 22(1):515–539, 1994.

[87] Peter M Robinson. Gaussian semiparametric estimation of long range dependence. *The Annals of statistics*, 10(2):1630–1661, 1995.

[88] Peter M Robinson. *Time series with long memory*. Oxford University Press, Oxford, 2003.

[89] M Röhrig and B Hardeweg. Efficient Farming Options for German Apple Growers Based On Stochastic Dominance Analysis. *Perspektiven für die Agrar-und Ernährungswirtschaft nach der Liberalisierung*, page 311, 2015.

[90] M Roth, C Ulardic, and J Trueb. Critical success factors for weather risk transfer solutions in the agricultural sector: A reinsurer's view. *Agricultural Finance Review*, 68(1):1–7, 2008.

[91] Walter Rudin. *Real and complex analysis, 3rd ed.* McGraw-Hill, Inc., New York, NY, USA, 1987.

[92] Anthony Saunders and Hugh Thomas. *Financial institutions management.* Irwin Boston, 1997.

[93] M. Schal. On Quadratic Cost Criteria for Option Hedging. *Mathematics of Operations Research*, 19(1):121–131, 1994.

[94] Zhiwei Shen and Martin Odening. Coping with systemic risk in index-based crop insurance. *Agricultural Economics (United Kingdom)*, 44(1):1–13, 2013.

[95] Katsumi Shimotsu. Exact Local Whittle Estimation of Fractional Integration With Unknown Mean and Time Trend. *Econometric Theory*, 26(02):501, 2010.

[96] Katsumi Shimotsu and Peter C. B. Phillips. Exact local Whittle estimation of fractional integration. *The Annals of Statistics*, 33(4):1890–1933, 2005.

[97] Jeremy Smith, Nick Taylor, and Sanjay Yadav. Comparing the bias and misspecification in ARFIMA models. *Journal of Time Series Analysis*, 18(5):507–527, 1997.

[98] Vincent H. Smith and Barry K. Goodwin. Crop Insurance, Moral Hazard and Agricultural Chemical Use. *American Journal of Agricultural Economics*, 78(2):428–438, 1996.

[99] R L. Snyder and J P D. Melo-Abreu. *Frost Protection: Fundamentals, Practice and Economics, 1.*, volume 1. FAO, 2005.

[100] Fallaw Sowell. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of econometrics*, 53(1):165–188, 1992.

[101] Joseph Stiglitz. Risk , Incentives and Insurance : The Pure Theory of Moral Hazard. *The Geneva Papers on Risk and Insurance*, 8(26):4–33, 1983.

[102] Joseph E. Stiglitz. Incentives and Risk Sharing in Sharecropping. *The Review of Economic Studies*, 41(2):219–255, 1974.

[103] Baojing Sun and G. Cornelis van Kooten. Financial Weather Options for Crop Production. Working papers, University of Victoria, Department of Economics, Resource Economics and Policy Analysis Research Group, feb 2014.

[104] Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov. Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35(6):2769–2794, 2007.

[105] Thorvald Nicolai Thiele. Om anvendelse af mindste kvadraters methode i nogle tilfælde, hvor en komplikation af visse slags uensartede tilfældige fejlkilder giver fejlene en 'systematisk'karakter. *Det Kongelige Danske Videnskabernes Selskabs Skrifter-Naturvidenskabelig og Mathematisk Afdeling*, pages 381–408, 1880.

[106] Calum G. Turvey. Weather derivatives for specific event risks in agriculture. *Review of Agricultural Economics*, 23(2):333–351, 2001.

[107] Calum G. Turvey. A note on scaled variance ratio estimation of the Hurst exponent with application to agricultural commodity prices. *Physica A: Statistical Mechanics and its Applications*, 377(1):155–165, 2007.

[108] Calum G. Turvey and Michael Norton. An internet-based tool for weather risk management. *Agricultural and Resource Economics Review*, 37(1):63–78, 2008.

[109] Raymond Venner and Steven C Blank. *Reducing Citrus Revenue Losses from Frost Damage*. Giannini Foundation information series. Giannini Foundation of Agricultural Economics, University of California, 1995.

[110] Shaun S. Wang. Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17(1):43–54, 1995.

[111] Shaun S. Wang. A class of distortion operators for pricing financial and insurance risks. *Journal of risk and insurance*, pages 15–36, 2000.

[112] RD Weaver and Taeho Kim. Designing crop insurance to manage moral hazard costs. *Zaragoza (Spain)*, 28:31, 2002.

[113] Peter Whitle. *Hypothesis testing in time series analysis*. PhD thesis, Uppsala University, 1951.

[114] Junjie Wu. Crop Insurance, Acreage Decisions, and Nonpoint-Source Pollution. *American Journal of Agricultural Economics*, 81(2):305–320, 1999.

[115] Xiang Yu; Sun Xiangyu and Zhong Funing. A Welfare Economics Analysis to the Subsidies of Agricultural Insurance. *Agricultural Economic Issues*, 2(6-13), 2008.

[116] Yuanchang Xu and Jiyu Jiang. The optimal boundary of political subsidies for agricultural insurance in welfare economic prospect. *Agriculture and Agricultural Science Procedia*, 1:163–169, 2010.

[117] Sheng Chao Zhuang, Chengguo Weng, Ken Seng Tan, and Hirbod Assa. Marginal Indemnification Function formulation for optimal reinsurance. *Insurance: Mathematics and Economics*, 67:65–76, 2016.

Appendix A

Appendix

A.1 Proof of Proposition 1

We have

$$
\begin{aligned}
&\mathbb{P}\left(D(T_I)1_{\{T_{I-k}\leq T_c\}} > x^\eta\right) \\
&= \mathbb{P}(D(T_I) > x^\eta \text{ and } T_{I-k} \leq T_c) \\
&= \mathbb{P}\left(\left(\frac{T_I - T_c}{T_t - T_c}\right)^\eta > x^\eta \text{ and } T_0 + (T_I - T_0)\sqrt{\frac{I-k}{I}} \leq T_c \text{ and } T_t \leq T_I \leq T_c\right) \\
&= \mathbb{P}((T_c - T_t)x < (T_0 - T_I) - (T_0 - T_c) \\
&\qquad \text{and } T_0 - T_c \leq (T_0 - T_I)\sqrt{\frac{I-k}{I}} \\
&\text{and } T_0 - T_c \leq T_0 - T_I \leq (T_0 - T_c) + (T_c - T_t)) \\
&= \mathbb{P}\left(bx < N - M \text{ and } M \leq BN \text{ and } M \leq N \leq M + b\right) \\
&= \mathbb{P}\left(bx + M < N \text{ and } \frac{M}{B} \leq N \leq M + b\right) \\
&= \int_{-\infty}^{\infty} \mathbb{P}\left(bx + z < N \text{ and } \frac{z}{B} \leq N \leq z + b \mid M = z\right) f_M(z)\, dz \\
&= \int_{-\infty}^{z_1} \mathbb{P}\left(bx + z < N \text{ and } \frac{z}{B} \leq N \leq z + b\right) f_M(z)\, dz \\
&= \int_{-\infty}^{z_x} \mathbb{P}\left(bx + z < N \leq z + b\right) f_M(z)\, dz + \int_{z_x}^{z_1} \mathbb{P}\left(\frac{z}{B} \leq N \leq z + b\right) f_M(z)\, dz \\
&= \int_{-\infty}^{z_x} \left(\bar{F}_N(bx + z) - \bar{F}_N(b + z)\right) f_M(z)\, dz \\
&+ \int_{z_x}^{z_1} \left(\bar{F}_N\left(\frac{z}{B}\right) - \bar{F}_N(b + z)\right) f_M(z)\, dz \\
&= \int_{-\infty}^{z_x} \bar{F}_N(bx + z) f_M(z)\, dz + \int_{z_x}^{+\infty} \bar{F}_N\left(\frac{z}{B}\right) f_M(z)\, dz \\
&- \left(\int_{-\infty}^{z_1} \bar{F}_N(b + z) f_M(z)\, dz + \int_{z_1}^{+\infty} \bar{F}_N\left(\frac{z}{B}\right) f_M(z)\, dz\right) \\
&= \mathcal{H}(x) - \mathcal{H}(1),
\end{aligned}
$$

where $z_x = \frac{Bbx}{1-B}$. Note that z_x is the point at which two line $y = bx + z$ and $y = \frac{z}{B}$ intersect, and also note that since $B < 1$ before z_x, $bx+z > \frac{z}{B}$ and after z_x, $bx+z < \frac{z}{B}$.

A.2 Proof of Proposition 2

Note that the CDF of the exponential distribution of $N \sim \exp(\lambda)$ is equal to $F_N(x) = 1-\exp(-\lambda x)$ and the PDF of $M \sim \mathcal{N}(\mu_M, \sigma_M)$ is equal to $f_M(x) = \frac{1}{\sqrt{2\pi}\sigma_M} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_M}{\sigma_M}\right)^2\right)$.

Now using them in (2.2.5), we have

$$\begin{aligned}\int_{-\infty}^{z_x} \bar{F}_N\left(bx+z\right) f_M\left(z\right) dz &= \int_{-\infty}^{z_x} \left(1-F_N\left(bx+z\right)\right) f_M\left(z\right) dz \\ &= \int_{-\infty}^{z_x} \exp\left(-\lambda\left(bx+z\right)\right) \frac{1}{\sqrt{2\pi}\sigma_M} \exp\left(-\frac{1}{2}\left(\frac{z-\mu_M}{\sigma_M}\right)^2\right) dz \\ &= \exp\left(\frac{\lambda^2\sigma_M^2}{2}-\lambda\left(bx+\mu_M\right)\right) \times \\ &\int_{-\infty}^{z_x} \frac{1}{\sqrt{2\pi}\sigma_M} \exp\left(-\frac{1}{2}\left(\frac{z_x-(\mu_M-\lambda\sigma_M^2)}{\sigma_M}\right)^2\right) dz \\ &= \exp\left(\frac{\lambda^2\sigma_M^2}{2}-\lambda\left(bx+\mu_M\right)\right) \Phi\left(\frac{z_x-(\mu_M-\lambda\sigma_M^2)}{\sigma_M}\right).\end{aligned}$$

On the other hand,

$$\begin{aligned}\int_{z_x}^{\infty} \bar{F}_N\left(\frac{z}{B}\right) f_M\left(z\right) dz &= \int_{z_x}^{\infty} \left(1-F_N\left(\frac{z}{B}\right)\right) f_M\left(z\right) dz \\ &= \int_{z_x}^{\infty} \exp\left(-\lambda\frac{z}{B}\right) \frac{1}{\sqrt{2\pi}\sigma_M} \exp\left(-\frac{1}{2}\left(\frac{z-\mu_M}{\sigma_M}\right)^2\right) dz \\ &= \exp\left(\frac{\lambda^2\sigma_M^2}{2B^2}-\frac{\lambda\mu_M}{B}\right) \left(1-\Phi\left(\frac{z-\left(\mu_M-\frac{\lambda\sigma_M^2}{B}\right)}{\sigma_M}\right)\right) \\ &= \exp\left(\frac{\lambda^2\sigma_M^2}{2B^2}-\frac{\lambda\mu_M}{B}\right) \Phi\left(\frac{\mu_M-\frac{\lambda\sigma_M^2}{B}-z_x}{\sigma_M}\right).\end{aligned}$$

Therefore, in this case we get

$$\begin{aligned}\mathcal{H}\left(x\right) = &\exp\left(\frac{\lambda^2\sigma_M^2}{2}-\lambda\left(bx+\mu_M\right)\right) \Phi\left(\frac{z_x-(\mu_M-\lambda\sigma_M^2)}{\sigma_M}\right) \\ &+ \exp\left(\frac{\lambda^2\sigma_M^2}{2B^2}-\frac{\lambda\mu_M}{B}\right) \Phi\left(\frac{\mu_M-\frac{\lambda\sigma_M^2}{B}-z_x}{\sigma_M}\right).\end{aligned}$$

A.3 Proof of Proposition 3

Note that the exponential distribution CDF of $N \sim \exp(\lambda)$ is equal to $F_N(x) = 1-\exp(-\lambda x)$ and the PDF of $M \sim \exp\left(\omega\right)$ is equal to $f_M(x) = \omega\exp\left(-\omega x\right)$. Now

using them in Eq.(2.2.5), we have

$$\begin{aligned}\int_{-\infty}^{z_x} \bar{F}_N\left(bx+z\right) f_M\left(z\right) dz &= \int_0^{z_x} \left(1-F_N\left(bx+z\right)\right) f_M\left(z\right) dz \\ &= \int_0^{z_x} \exp\left(-\lambda\left(bx+z\right)\right) \omega \exp\left(-\omega z\right) dz \\ &= \omega \exp\left(-\lambda bx\right) \int_0^{z_x} \exp\left(-\left(\lambda+\omega\right) z\right) dz \\ &= \frac{\omega \exp\left(-\lambda bx\right)}{\lambda+\omega} \left(1-\exp\left(-\left(\lambda+\omega\right) z_x\right)\right).\end{aligned}$$

On the other hand,

$$\begin{aligned}\int_{z_x}^{\infty} \bar{F}_N\left(\frac{z}{B}\right) f_M\left(z\right) dz &= \int_{z_x}^{\infty} \left(1-F_N\left(\frac{z}{B}\right)\right) f_M\left(z\right) dz \\ &= \int_{z_x}^{\infty} \exp\left(-\lambda\frac{z}{B}\right) \omega \exp\left(-\omega z\right) dz \\ &= \omega \int_{z_x}^{\infty} \exp\left(-\left(\frac{\lambda}{B}+\omega\right) z\right) dz \\ &= \frac{\omega}{\frac{\lambda}{B}+\omega} \exp\left(-\left(\frac{\lambda}{B}+\omega\right) z_x\right).\end{aligned}$$

Therefore, in this case we get

$$\begin{aligned}\mathcal{H}\left(x\right) = &\frac{\omega \exp\left(-\lambda bx\right)}{\lambda+\omega} \left(1-\exp\left(-\left(\lambda+\omega\right) z_x\right)\right) \\ &+ \frac{\omega}{\frac{\lambda}{B}+\omega} \exp\left(-\left(\frac{\lambda}{B}+\omega\right) z_x\right).\end{aligned}$$

A.4 Tables

The following are tables mentioned in the Chapter 2 and Chapter 3.

Commodity	Hurst	ARFIMA Model	AICBIC
S&P GSCI Lean Hogs Spot Index	0.8248 (0.7109, 0.8253)	2,1,2	-127670
LH Hogs, Lean Future CME	0.7624 (0.6738, 0.7848)	0,1,4	-78178
LB Lumber CME	0.8471 (0.7157, 0.8546)	2,1,1	-74520
LC Cattle Live CME	0.7997 (0.5910, 0.9880)	1,1,1	-114781
CME Lean Hogs Future	0.7618 (0.6744, 0.7943)	0,1,4	-78181
CBOT Oats Future	0.7497 (0.6780, 0.7840)	4,1,3	-127367
CME Live Cattle	0.7864 (0.5910, 0.9880)	5,1,2	-114789
NYBOT Or juice Future	0.7775 (0.6620, 0.7842)	6,1,3	-128888
CBOT Soybean Future	0.8546 (0.7080, 0.8633)	4,1,0	-143437
CBOT Wheat Future	0.8710 (0.7526, 0.8837)	1,1,1	-129942
CBOT Corn Future	0.7007 (0.6279, 0.8726)	5,1,1	-136744
CT Cotton NYBOT	0.8371 (0.7204, 0.8379)	3,1,1	-115284
NYBOT Coffee Future	0.7853 (0.7038, 0.8048)	0,1,2	-114269
WCE Canola Future	0.7566 (0.6820, 0.7806)	4,1,4	-111777
NYBOT Sugar Future	0.8636 (0.7506, 0.8659)	1,1,1	-117249
NYBOT Cocoa Future	0.8569 (0.7279, 0.8608)	2,1,2	-129917
CBOT Soybean Oil Future	0.8227 (0.7080, 0.8633)	2,1,2	-128342
FC Cattle Feeder CME	0.7877 (0.6969, 0.8054)	1,1,1	-94546
CME Feeder Cattle INDEX	0.6678 (0.5881, 0.7206)	6,1,6	-78989

Table A.4.1: Hurst exponent with $5\% - 95\%$ estimation interval from the absolute log return price

Commodity	Hurst	ARFIMA Model	AICBIC
S&P GSCI Lean Hogs Spot Index	0.3927 (0.3915 0.5308)	6,1,6	100120
LH Hogs, Lean Future CME	0.3998 (0.3989 0.5015)	1,1,1	57595
LB Lumber CME	0.4088 (0.4050 0.4966)	1,1,2	94617
LC Cattle Live CME	0.4109 (0.4013 0.5640)	0,1,4	55838
CME Lean Hogs Future	0.4141 (0.4089 0.4915)	1,1,1	49318
CBOT Oats Future	0.4432 (0.4229 0.5549)	5,1,4	134573
CME Live Cattle	0.4540 (0.4092 0.49338)	4,1,4	43328
NYBOT Or juice Future	0.4583 (0.4289 0.5353)	3,1,1	100989
CBOT Soybean Future	0.4748 (0.4484 0.5784)	2,1,3	185141
CBOT Wheat Future	0.4775 (0.4592 0.5390)	2,1,2	160399
CBOT Corn Future	0.4815 (0.4516 0.5384)	5,1,5	147804
CT Cotton NYBOT	0.4982 (0.4729 0.5921)	3,1,6	92313
NYBOT Coffee Future	0.4989 (0.4728 0.5604)	5,1,2	112961
WCE Canola Future	0.5062 (0.4776 0.5705)	4,1,4	103463
NYBOT Sugar Future	0.5074 (0.4796 0.5799)	5,1,4	23952
NYBOT Cocoa Future	0.5102 (0.4643 0.6062)	6,1,3	229260
CBOT Soybean Oil Future	0.5325 (0.4893 0.6552)	6,1,6	27593
FC Cattle Feeder CME	0.5887 (0.5588 0.6557)	2,1,2	38578
CME Feeder Cattle INDEX	0.6259 (0.5960 0.6962)	5,1,2	20322

Table A.4.2: Hurst exponent with $5\% - 95\%$ estimation interval from the increment price

		AR	FI	MA	Average Err.
	Real	**0.5**	**0.3**		
ARFIMA	**RS**	0.5535	0.2548		0.0350
0.5,0.3,0	**AV**	0.5654	0.2431		0.0433
	OT	0.5575	0.2508		0.0378
	Average	0.5591	0.2492		0.0390
	Real	**0.5**	**-0.3**		
ARFIMA	**RS**	0.3545	-0.1599		0.1010
0.5,-0.3,0	**AV**	0.5040	-0.3122		0.0064
	OT	0.4651	-0.2735		0.0219
	Average	0.4401	-0.2483		0.0395
	Real		**0.2**	**0.2**	
ARFIMA	**RS**		0.1802	0.2125	0.0117
0,0.2,0.2	**AV**		0.1386	0.2484	0.0391
	OT		0.1639	0.2266	0.0224
	Average		0.1607	0.2293	0.0245
	Real		**-0.2**	**0.2**	
ARFIMA	**RS**		-0.1276	0.1239	0.0525
0,-0.2,0.2	**AV**		-0.2344	0.2182	0.0195
	OT		-0.2262	0.2110	0.0142
	Average		-0.1964	0.1849	0.0078
	Real	**0.5**	**0.1**	**0.2**	
ARFIMA	**RS**	0.4107	0.1828	0.2141	0.0409
0.5,0.1,0.2	**AV**	0.4702	0.1322	0.2055	0.0148
	OT	0.4742	0.1288	0.2050	0.0130
	Average	0.4524	0.1475	0.2078	0.0226
	Real	**0.5**	**-0.1**	**0.2**	
ARFIMA	**RS**	0.3936	-0.0017	0.2271	0.0491
0.5,-0.1,0.2	**AV**	0.5054	-0.0982	0.2125	0.0046
	OT	0.5209	-0.1120	0.2112	0.0089
	Average	0.4740	-0.0706	0.2157	0.0141

Table A.4.3: Accuracy test

Table A.4.4: Table of Hurst estimation methods stabilities

d	Model	ARFIMA	0,d,0	10000	ARFIMA	0.5,d,0	10000	ARFIMA	0,d,0.3	10000	ARFIMA	0.5,d,0.2	10000	fBM	d+1	10000
		Method			Method			Method			Method			Method	First difference	
	Stats	RS	AV	OT	RS	AV	OT	RS	AV	OT	RS	AV	OT	RS	AV	OT
-0.5	Mean	-0.279968	-0.423363	-0.402521	-0.212742	-0.387203	-0.320102	-0.257004	-0.411348	-0.372613	-0.201618	-0.382799	-0.304920	-0.305253	-0.437374	-0.430768
	Std.	0.008970	0.014603	0.029589	0.009791	0.015963	0.029862	0.009160	0.015372	0.029168	0.009983	0.016310	0.028688	0.008373	0.013014	0.029327
-0.4	Mean	-0.234204	-0.363694	-0.343588	-0.163616	-0.324702	-0.257297	-0.210269	-0.350438	-0.313828	-0.153634	-0.320417	-0.242862	-0.256719	-0.376678	-0.368953
	Std.	0.010218	0.017184	0.028535	0.011385	0.018571	0.028536	0.010834	0.017762	0.028663	0.011467	0.018729	0.028907	0.009945	0.016031	0.028848
-0.3	Mean	-0.177702	-0.287244	-0.270438	-0.107324	-0.250307	-0.185323	-0.154015	-0.275308	-0.240844	-0.098468	-0.247510	-0.172931	-0.195200	-0.296527	-0.290274
	Std.	0.012168	0.020113	0.028096	0.013218	0.021136	0.027981	0.012522	0.020402	0.028473	0.013450	0.021532	0.027593	0.011772	0.019179	0.027641
-0.2	Mean	-0.112445	-0.199786	-0.188175	-0.045193	-0.168250	-0.107885	-0.089908	-0.190219	-0.159039	-0.037227	-0.165777	-0.095980	-0.123552	-0.205396	-0.200271
	Std.	0.014491	0.023289	0.027763	0.015269	0.023431	0.026714	0.014589	0.023308	0.027473	0.015331	0.023805	0.027815	0.014262	0.022976	0.027413
-0.1	Mean	-0.040111	-0.106369	-0.098209	0.021633	-0.081374	-0.026382	-0.019612	-0.099609	-0.073008	0.028611	-0.078497	-0.015416	-0.044546	-0.108198	-0.103433
	Std.	0.016888	0.026127	0.027479	0.017406	0.026289	0.027675	0.017225	0.026336	0.027375	0.017094	0.025912	0.027151	0.017106	0.026079	0.027407
0	Mean	0.037691	-0.011074	-0.006004	0.091437	0.007469	0.056708	0.054800	-0.007138	0.015244	0.096370	0.009498	0.065634	0.037607	-0.011592	-0.006626
	Std.	0.019328	0.029161	0.028472	0.019481	0.028474	0.027994	0.019680	0.029313	0.028179	0.019757	0.028965	0.028084	0.019377	0.029039	0.028166
0.1	Mean	0.117959	0.082196	0.085353	0.162628	0.097227	0.140054	0.131476	0.085973	0.104074	0.165103	0.098236	0.145844	0.122339	0.084286	0.090286
	Std.	0.021879	0.032085	0.029640	0.021386	0.031054	0.028563	0.022057	0.032305	0.029815	0.021682	0.031927	0.029858	0.021638	0.031865	0.029290
0.2	Mean	0.199294	0.172899	0.175925	0.230719	0.182208	0.218204	0.208509	0.175302	0.190802	0.231876	0.182470	0.222715	0.206664	0.175834	0.184005
	Std.	0.023508	0.034515	0.030617	0.023460	0.034654	0.031087	0.023409	0.034566	0.031175	0.023128	0.034979	0.031003	0.024103	0.034468	0.030930
0.3	Mean	0.276544	0.254610	0.259019	0.294540	0.261357	0.291414	0.281279	0.256042	0.270628	0.294510	0.261887	0.294296	0.287409	0.260532	0.271124
	Std.	0.025167	0.036576	0.032294	0.024219	0.036274	0.031515	0.024826	0.037494	0.032784	0.024274	0.036355	0.031631	0.025743	0.035364	0.031711
0.4	Mean	0.347163	0.327971	0.334992	0.351028	0.331667	0.355215	0.345771	0.327742	0.341323	0.349404	0.331373	0.356389	0.358304	0.332396	0.346412
	Std.	0.025945	0.037171	0.032610	0.023931	0.036567	0.031289	0.025397	0.037093	0.032332	0.023891	0.036176	0.030782	0.026420	0.034008	0.030523
0.5	Mean	0.403169	0.384160	0.393219	0.395565	0.386959	0.404183	0.398019	0.385324	0.397667	0.393529	0.387018	0.404992	0.413473	0.386695	0.402586
	Std.	0.024060	0.034560	0.029745	0.022737	0.034371	0.028729	0.023627	0.034818	0.029743	0.023060	0.034661	0.028778	0.024246	0.029042	0.026283
Average Err.	**Mean**	**0.031039**	**0.015544**	**0.016221**	**0.044614**	**0.018528**	**0.028196**	**0.035524**	**0.016309**	**0.019214**	**0.046519**	**0.019023**	**0.030796**	**0.027227**	**0.014231**	**0.012679**
Average	**Std.**	**0.018420**	**0.027762**	**0.029531**	**0.018389**	**0.027889**	**0.029086**	**0.018484**	**0.028070**	**0.029562**	**0.018465**	**0.028123**	**0.029117**	**0.018453**	**0.026460**	**0.028867**

	Models	ARFIMA		ARFIMA		ARFIMA		
	p,d,q	(0.5,0.3,0)	(0.5,-0.3,0)	(0,0.2,0.2)	(0,-0.2,0.2)	(0.5,0.1,0.2)	(0.5,-0.1,0.2)	Average
Method	Steps	0.3	-0.3	0.2	-0.2	0.1	-0.1	Error
RS	1	0.282508	-0.120719	0.188014	-0.117429	0.212298	0.044687	**0.045017**
	2	0.259761	-0.153673	0.180409	-0.126846	0.187427	0.007140	**0.036472**
	3	0.255753	-0.158888	0.180165	-0.127535	0.183607	0.000085	**0.035167**
	4	0.254986	-0.159766	0.180157	-0.127585	0.182974	-0.001351	**0.034939**
	5	0.254837	-0.159916	0.180157	-0.127588	0.182868	-0.001648	**0.034897**
	10	0.254801	-0.159946	0.180157	-0.127589	0.182847	-0.001725	**0.034887**
	15	0.254801	-0.159946	0.180157	-0.127589	0.182847	-0.001725	**0.034887**
	20	0.254801	-0.159946	0.180157	-0.127589	0.182847	-0.001725	**0.034887**
	25	0.254801	-0.159946	0.180157	-0.127589	0.182847	-0.001725	**0.034887**
	30	0.254801	-0.159946	0.180157	-0.127589	0.182847	-0.001725	**0.034887**
	45	0.254801	-0.159946	0.180157	-0.127589	0.182847	-0.001725	**0.034887**
AV	1	0.253092	-0.273630	0.140896	-0.226125	0.142337	-0.072987	**0.016325**
	2	0.243741	-0.305453	0.138629	-0.234143	0.132896	-0.094750	**0.016018**
	3	0.243174	-0.310923	0.138612	-0.234419	0.132289	-0.097701	**0.016115**
	4	0.243139	-0.311952	0.138612	-0.234429	0.132249	-0.098133	**0.016136**
	5	0.243137	-0.312149	0.138612	-0.234429	0.132246	-0.098197	**0.016140**
	10	0.243136	-0.312196	0.138612	-0.234429	0.132246	-0.098208	**0.016141**
	15	0.243136	-0.312196	0.138612	-0.234429	0.132246	-0.098208	**0.016141**
	20	0.243136	-0.312196	0.138612	-0.234429	0.132246	-0.098208	**0.016141**
	25	0.243136	-0.312196	0.138612	-0.234429	0.132246	-0.098208	**0.016141**
	30	0.243136	-0.312196	0.138612	-0.234429	0.132246	-0.098208	**0.016141**
	45	0.243136	-0.312196	0.138612	-0.234429	0.132246	-0.098208	**0.016141**
OT	1	0.293221	-0.193101	0.176548	-0.204196	0.182024	-0.025177	**0.026017**
	2	0.258793	-0.252295	0.164501	-0.224179	0.139805	-0.087479	**0.014491**
	3	0.252460	-0.267381	0.163916	-0.226019	0.131268	-0.104351	**0.013228**
	4	0.251173	-0.271671	0.163887	-0.226188	0.129355	-0.109562	**0.013049**
	5	0.250907	-0.272926	0.163886	-0.226204	0.128917	-0.111232	**0.013013**
	10	0.250837	-0.273451	0.163886	-0.226206	0.128787	-0.112031	**0.013002**
	15	0.250837	-0.273452	0.163886	-0.226206	0.128787	-0.112034	**0.013002**
	20	0.250837	-0.273452	0.163886	-0.226206	0.128787	-0.112034	**0.013002**
	25	0.250837	-0.273452	0.163886	-0.226206	0.128787	-0.112034	**0.013002**
	30	0.250837	-0.273452	0.163886	-0.226206	0.128787	-0.112034	**0.013002**
	45	0.250837	-0.273452	0.163886	-0.226206	0.128787	-0.112034	**0.013002**
Average	1	0.276274	-0.195817	0.168486	-0.182583	0.178887	-0.017826	**0.026713**
	2	0.253173	-0.238164	0.160906	-0.195481	0.152101	-0.059895	**0.018172**
	3	0.249850	-0.246228	0.160683	-0.196336	0.148224	-0.068243	**0.016915**
	4	0.249335	-0.247898	0.160676	-0.196393	0.147621	-0.070068	**0.016671**
	5	0.249254	-0.248250	0.160676	-0.196396	0.147526	-0.070474	**0.016619**
	10	0.249239	-0.248344	0.160676	-0.196397	0.147508	-0.070590	**0.016605**
	15	0.249239	-0.248344	0.160676	-0.196397	0.147508	-0.070590	**0.016605**
	20	0.249239	-0.248344	0.160676	-0.196397	0.147508	-0.070590	**0.016605**
	25	0.249239	-0.248344	0.160676	-0.196397	0.147508	-0.070590	**0.016605**
	30	0.249239	-0.248344	0.160676	-0.196397	0.147508	-0.070590	**0.016605**
	45	0.249239	-0.248344	0.160676	-0.196397	0.147508	-0.070590	**0.016605**
	p,d,q	0.5,0.3,0	0.5,-0.3,0	0,0.2,0.2	0,-0.2,0.2	0.5,0.1,0.2	0.5,-0.1,0.2	

Table A.4.5: Table of the estimation convergence

Product name	Days	Mean	Variance	Skweness	Kurtosis	KPSS	ADF
BRENT CRUDE OIL INDEX	4383	62.83	1213	0.2473	1.7393	0.01	0.57
CBOT CORN FUTURE	11356	295.51	16622	1.8275	6.4470	0.01	0.40
CBOT OATS FUTURE	11452	180.07	6459	1.2847	4.2692	0.01	0.29
CBOT SOYBN FUTURE	11457	713.43	79087	1.3174	4.4276	0.01	0.44
CBOT SOYBN OIL FUTURE	10191	26.94	110	1.5061	4.8304	0.01	0.27
CBOT WHEAT FUTURE	10992	403.76	23451	1.5125	5.5088	0.01	0.35
CL CRUDE OIL NYMEX	6624	46.11	979	0.8023	2.2901	0.01	0.41
CME FEEDER CATTLE INDEX	4805	109.85	1518	1.5069	5.1426	0.01	1.00
CT COTTON NYBOT	10080	68.15	368	2.6049	16.5470	0.01	0.34
FC CATTLE FEEDER CME	6446	102.04	1283	1.7922	6.2853	0.01	0.99
GENERIC 1ST FUTURE GOL	10093	534.29	153494	1.6741	4.7814	0.01	0.84
HG COPPER NYMEX	6685	181.52	13005	0.7429	1.9573	0.01	0.50
HO HEATING OIL NYMEX	7296	125.22	8782	0.9644	2.5198	0.01	0.46
LA ALUMINUM FUTURE	4486	1874.19	217237	0.7035	2.5746	0.01	0.47
LA CRUDE OIL SPOT	6163	41.47	1024	0.8416	2.2766	0.01	0.46
LB LUMBER CME	6646	282.54	4487	0.2529	2.4577	0.01	0.39
LC CATTLE LIVE CME	8347	81.99	539	1.6916	5.4931	0.01	0.88
HOGS, LEAN FUTURE CME	7385	61.15	301	1.0170	4.1324	0.01	0.43
LME 3M COPPER FUTURE	7411	3777.42	6095279	0.8835	2.1950	0.01	0.53
LME ALUMINUM SPOT	7034	1785.20	224536	0.9541	3.6849	0.01	0.35
LME COPPER SPOT	7395	3805.70	6104507	0.8680	2.1900	0.01	0.52
LME PL INDEX	5873	884.98	248848	0.5268	1.9135	0.01	0.58
NYBOT COCOA FUTURE	11209	1825.89	674508	0.7549	3.2154	0.01	0.48
NYBOT COFFEE FUTURE	10648	125.17	2660	0.7929	3.6491	0.01	0.21
NYBOT OR JUICE FUTRUE	11348	109.78	1521	0.2347	2.4749	0.01	0.36
NYBOT SUGAR FUTURE	11312	12.01	48	2.1007	9.2338	0.01	0.13
NYMEX CRUDE FUTURE	8094	41.83	892	1.0768	2.8560	0.01	0.35
PA PALLADIUM NYMEX	7269	326.72	55634	0.9770	2.6874	0.01	0.57
PL PLATINUM NYMEX	7941	107.68	1542	0.8977	3.2269	0.01	0.55
GSCI COPPER INDEX SPOT	7354	807.26	227179	0.8662	2.3940	0.01	0.50
GSCI LEAN HOGS SPOT INDEX	9730	230.98	27862	1.2550	3.0455	0.01	0.36
GSCI PALLADIUM INDEX ER	9983	98.53	376	0.8783	5.6554	0.01	0.70
SI SILVER NYMEX	1723	415.07	17772	-0.9263	2.5756	0.01	0.21
WCE CANOLA FUTRUE	10123	9.74	61	1.9738	6.6083	0.01	0.59
WTI CUSING CRUDE SPOT	8421	377.23	9224	0.9768	3.6844	0.01	0.35
XAG SILVER SPOT	8020	41.96	895	1.0688	2.8372	0.01	0.15
XAU GOLD SPOT	11520	9.06	60	2.0231	6.9884	0.01	0.85
XPD PALLADIUM SPOT	10521	527.46	156339	1.6412	4.7355	0.01	0.60
XPT PLATINUM SPOT	5572	398.22	54108	0.6520	2.1052	0.01	0.56
PH GOLD&SILVER INDEX	7441	813.35	228838	0.8320	2.3314	0.01	0.23

Table A.4.6: Data statistical properties

Product name	Diff KPSS	Diff ADF	Log KPSS	Log ADF
BRENT CRUDE OIL INDEX	0.0497	0.0010	0.1000	0.0010
CBOT CORN FUTURE	0.1000	0.0010	0.1000	0.0010
CBOT OATS FUTURE	0.1000	0.0010	0.1000	0.0010
CBOT SOYBN FUTURE	0.1000	0.0010	0.1000	0.0010
CBOT SOYBN OIL FUTURE	0.1000	0.0010	0.1000	0.0010
CBOT WHEAT FUTURE	0.1000	0.0010	0.1000	0.0010
CL CRUDE OIL NYMEX	0.1000	0.0010	0.1000	0.0010
CME FEEDER CATTLE INDEX	0.0162	0.0010	0.0345	0.0010
CT COTTON NYBOT	0.1000	0.0010	0.1000	0.0010
FC CATTLE FEEDER CME	0.1000	0.0010	0.1000	0.0010
GENERIC 1ST FUTURE GOL	0.1000	0.0010	0.1000	0.0010
HG COPPER NYMEX	0.1000	0.0010	0.1000	0.0010
HO HEATING OIL NYMEX	0.1000	0.0010	0.1000	0.0010
LA ALUMINUM FUTURE	0.1000	0.0010	0.1000	0.0010
LA CRUDE OIL SPOT	0.0825	0.0010	0.1000	0.0010
LB LUMBER CME	0.1000	0.0010	0.1000	0.0010
LC CATTLE LIVE CME	0.1000	0.0010	0.1000	0.0010
HOGS, LEAN FUTURE CME	0.1000	0.0010	0.1000	0.0010
LME 3M COPPER FUTURE	0.1000	0.0010	0.1000	0.0010
LME ALUMINUM SPOT	0.1000	0.0010	0.1000	0.0010
LME COPPER SPOT	0.1000	0.0010	0.1000	0.0010
LME PL INDEX	0.1000	0.0010	0.1000	0.0010
NYBOT COCOA FUTURE	0.1000	0.0010	0.1000	0.0010
NYBOT COFFEE FUTURE	0.1000	0.0010	0.1000	0.0010
NYBOT OR JUICE FUTRUE	0.1000	0.0010	0.1000	0.0010
NYBOT SUGAR FUTURE	0.1000	0.0010	0.1000	0.0010
NYMEX CRUDE FUTURE	0.1000	0.0010	0.1000	0.0010
PA PALLADIUM NYMEX	0.1000	0.0010	0.1000	0.0010
PL PLATINUM NYMEX	0.1000	0.0010	0.1000	0.0010
GSCI COPPER INDEX SPOT	0.1000	0.0010	0.1000	0.0010
GSCI LEAN HOGS SPOT INDEX	0.1000	0.0010	0.1000	0.0010
GSCI PALLADIUM INDEX ER	0.1000	0.0010	0.1000	0.0010
SI SILVER NYMEX	0.1000	0.0010	0.1000	0.0010
WCE CANOLA FUTRUE	0.1000	0.0010	0.1000	0.0010
WTI CUSING CRUDE SPOT	0.1000	0.0010	0.1000	0.0010
XAG SILVER SPOT	0.1000	0.0010	0.1000	0.0010
XAU GOLD SPOT	0.1000	0.0010	0.0100	0.0010
XPD PALLADIUM SPOT	0.1000	0.0010	0.1000	0.0010
XPT PLATINUM SPOT	0.1000	0.0010	0.0867	0.0010
PH GOLD&SILVER INDEX	0.1000	0.0010	0.1000	0.0010

Table A.4.7: Stationary test for the increments and log difference data

Product	Hurst (Diff)	Hurst (Log)	RS (Diff)	RS (Log)
GSCI LEAN HOGS SPOT INDEX	0.3571	0.8248	0.5267	0.6102
HOGS, LEAN FUTURE CME	0.4141	0.7618	0.5170	0.6938
LB LUMBER CME	0.4403	0.8436	0.5168	0.7831
CBOT OATS FUTURE	0.4432	0.7433	0.5111	0.7714
LC CATTLE LIVE CME	0.4540	0.7863	0.4911	0.7664
PH GOLD&SILVER INDEX	0.4574	0.7307	0.5720	0.8187
NYBOT OR JUICE FUTRUE	0.4583	0.7747	0.4956	0.7481
CBOT SOYBN FUTURE	0.4748	0.8460	0.5237	0.8569
CBOT WHEAT FUTURE	0.4775	0.8703	0.5265	0.8206
CBOT CORN FUTURE	0.4815	0.7140	0.5225	0.8234
XAG SILVER SPOT	0.4870	0.8652	0.5713	0.8575
SI SILVER NYMEX	0.4932	0.8535	0.5409	0.7506
LME ALUMINUM SPOT	0.4937	0.8674	0.5599	0.8439
GSCI PALLADIUM INDEX ER	0.4972	0.6825	0.5840	0.8604
NYBOT COFFEE FUTURE	0.4989	0.7853	0.5472	0.7741
CT COTTON NYBOT	0.5001	0.8385	0.5607	0.7909
WCE CANOLA FUTRUE	0.5062	0.7566	0.5477	0.8458
NYBOT SUGAR FUTURE	0.5074	0.8636	0.5781	0.8276
NYBOT COCOA FUTURE	0.5102	0.8558	0.5232	0.8161
GSCI COPPER INDEX SPOT	0.5300	0.8726	0.5201	0.7907
CBOT SOYBN OIL FUTURE	0.5325	0.8225	0.5531	0.8265
LME 3M COPPER FUTURE	0.5377	0.8831	0.5549	0.8650
LME COPPER SPOT	0.5406	0.8488	0.5463	0.8664
HG COPPER NYMEX	0.5409	0.8473	0.5362	0.8190
HO HEATING OIL NYMEX	0.5412	0.8075	0.5542	0.8184
XPT PLATINUM SPOT	0.5487	0.8378	0.5066	0.7927
NYMEX CRUDE FUTURE	0.5512	0.8590	0.5470	0.8426
LA ALUMINUM FUTURE	0.5522	0.8488	0.5616	0.7610
PL PLATINUM NYMEX	0.5542	0.8440	0.5389	0.8776
CL CRUDE OIL NYMEX	0.5561	0.8287	0.5601	0.8280
LME PL INDEX	0.5590	0.8324	0.5667	0.8091
WTI CUSING CRUDE SPOT	0.5605	0.7799	0.5793	0.8755
LA CRUDE OIL SPOT	0.5676	0.8323	0.5815	0.8053
XPD PALLADIUM SPOT	0.5724	0.8036	0.5803	0.8157
XAU GOLD SPOT	0.5770	0.8346	0.5962	0.8081
PA PALLADIUM NYMEX	0.5812	0.8017	0.5967	0.7898
BRENT CRUDE OIL INDEX	0.5948	0.8693	0.6111	0.7835
FC CATTLE FEEDER CME	0.6031	0.7878	0.5740	0.7312
GENERIC 1ST FUTURE GOL	0.6192	0.8914	0.5745	0.8676
CME FEEDER CATTLE INDEX	0.6560	0.6680	0.6556	0.6883

Table A.4.8: Hurst exponent for increments and log difference for our algorithm and the RS method

Model	NoS	d	C	AR_1	AR_2	AR_3	AR_4	AR_5	AR_6	MA_1	MA_2	MA_3	MA_4	MA_5	MA_6	Var	AICBIC
ARMA(0_0)	9	-0.0124	0.0260													39.3506	147885
ARMA(0_1)	9	-0.0143	0.0263							0.0341						39.3128	147874
ARMA(0_2)	9	-0.0109	0.0258							0.0309	-0.0358					39.2620	147856
ARMA(0_3)	9	-0.0136	0.0261							0.0325	-0.0342	0.0215				39.2473	147859
ARMA(0_4)	9	-0.0151	0.0263							0.0340	-0.0337	0.0217	0.0099			39.2447	147869
ARMA(0_5)	9	-0.0169	0.0266							0.0360	-0.0331	0.0224	0.0101	0.0110		39.2409	147878
ARMA(0_6)	9	-0.0166	0.0265							0.0358	-0.0334	0.0224	0.0101	0.0108	-0.0019	39.2406	147889
ARMA(1_0)	9	-0.0143	0.0254	0.0317												39.3155	147876
ARMA(1_1)	9	-0.0130	0.0478	-0.8256						0.8571						39.2305	147838
ARMA(1_2)	9	-0.0132	0.0480	-0.8311						0.8644	0.0027					39.2304	147849
ARMA(1_3)	22	-0.0292	0.0061	0.7834						-0.7353	-0.0641	0.0441				39.2353	147864
ARMA(1_4)	22	-0.0296	0.0055	0.8034						-0.7549	-0.0649	0.0473	-0.0048			39.2346	147875
ARMA(1_5)	22	-0.0297	0.0060	0.7869						-0.7384	-0.0641	0.0469	-0.0072	0.0043		39.2342	147886
ARMA(1_6)	9	-0.0150	0.0477	-0.8046						0.8403	-0.0066	-0.0063	0.0287	0.0182	-0.0115	39.1930	147873
ARMA(2_0)	9	-0.0108	0.0260	0.0292	-0.0361											39.2647	147858
ARMA(2_1)	9	-0.0132	0.0479	-0.8279	0.0025					0.8611						39.2303	147849
ARMA(2_2)	9	-0.0133	0.0668	-1.1926	-0.2933					1.2288	0.3113					39.2294	147860
ARMA(2_3)	22	-0.0297	0.0062	0.6703	0.1094					-0.6217	-0.1678	0.0392				39.2345	147875
ARMA(2_4)	23	-0.0301	0.0098	-0.0548	0.7070					0.1043	-0.7304	-0.0037	0.0248			39.2043	147868
ARMA(2_5)	22	-0.0289	0.0123	-0.0918	0.6538					0.1406	-0.6766	-0.0079	0.0269	0.0115		39.2006	147877
ARMA(2_6)	22	-0.0286	0.0091	-0.0135	0.6889					0.0624	-0.7160	-0.0077	0.0338	0.0080	-0.0148	39.1942	147885
ARMA(3_0)	9	-0.0139	0.0256	0.0331	-0.0353	0.0236										39.2468	147859
ARMA(3_1)	21	-0.0288	0.0064	0.7886	-0.0632	0.0451				-0.7412						39.2369	147865
ARMA(3_2)	23	-0.0298	0.0065	0.6322	0.0993	0.0384				-0.5835	-0.1556					39.2345	147875
ARMA(3_3)	9	-0.0236	0.0019	1.1124	0.5938	-0.7739				-1.0759	-0.6573	0.8086				39.1138	147816
ARMA(3_4)	29	-0.0236	0.0087	0.1848	0.6859	-0.1890				-0.1416	-0.7240	0.1919	0.0271			39.2014	147878
ARMA(3_5)	22	-0.0292	0.0217	-0.9283	0.5713	0.5910				0.9778	-0.5527	-0.6202	0.0225	0.0387		39.1786	147876
ARMA(3_6)	23	-0.0293	0.0152	-0.6464	0.6341	0.4732				0.6962	-0.6293	-0.4959	0.0296	0.0269	-0.0177	39.1753	147886
ARMA(4_0)	9	-0.0153	0.0255	0.0344	-0.0343	0.0239	0.0088									39.2448	147869
ARMA(4_1)	23	-0.0296	0.0054	0.8298	-0.0658	0.0506	-0.0076			-0.7814						39.2353	147875
ARMA(4_2)	23	-0.0301	0.0101	-0.0536	0.6723	-0.0027	0.0255			0.1036	-0.6963					39.2033	147868
ARMA(4_3)	22	-0.0295	0.0091	0.1516	0.6568	-0.1591	0.0272			-0.1023	-0.6914	0.1602				39.2018	147878
ARMA(4_4)	26	-0.0295	0.0210	-0.7063	-0.2627	0.5370	0.7013			0.7521	0.2792	-0.4971	-0.7063			39.1409	147854
ARMA(4_5)	20	-0.0293	0.0252	-1.1929	0.0979	0.8248	0.3712			1.2426	-0.0651	-0.8331	-0.3598	0.0338		39.1664	147880
ARMA(4_6)	22	-0.0298	0.0356	-1.5318	-0.1915	0.9762	0.5314			1.5821	0.2416	-0.9819	-0.5301	0.0474	0.0205	39.1627	147890
ARMA(5_0)	9	-0.0172	0.0254	0.0362	-0.0336	0.0249	0.0089	0.0098								39.2419	147879
ARMA(5_1)	22	-0.0298	0.0060	0.8062	-0.0649	0.0502	-0.0102	0.0049		-0.7577						39.2346	147886
ARMA(5_2)	9	-0.0155	0.0786	-1.4302	-0.5348	-0.0068	0.0253	0.0330		1.4662	0.5520					39.1842	147868
ARMA(5_3)	23	-0.0303	0.0232	-1.0292	0.5802	0.6413	0.0224	0.0340		1.0799	-0.5553	-0.6700				39.1788	147876
ARMA(5_4)	24	-0.0300	0.0260	-1.2257	0.0382	0.8103	0.4221	0.0322		1.2761	-0.0024	-0.8154	-0.4104			39.1667	147881
ARMA(5_5)	9	-0.0218	0.0049	0.3750	0.4367	0.7561	0.0398	-0.7882		-0.3414	-0.4740	-0.7267	-0.0720	0.8131		39.0154	147804
ARMA(5_6)	12	-0.0203	0.0055	0.3384	0.4537	0.7355	0.0723	-0.8035		-0.3001	-0.4920	-0.7167	-0.1040	0.8262	0.0142	39.0127	147814
ARMA(6_0)	9	-0.0178	0.0253	0.0368	-0.0333	0.0251	0.0092	0.0098	0.0029							39.2419	147890
ARMA(6_1)	9	-0.0152	0.0473	-0.8129	-0.0054	-0.0047	0.0283	0.0172	-0.0103	0.8482						39.1951	147874
ARMA(6_2)	9	-0.0150	0.0763	-1.3785	-0.4963	-0.0071	0.0243	0.0286	-0.0039	1.4141	0.5108					39.1835	147879
ARMA(6_3)	23	-0.0294	0.0161	-0.7189	0.6249	0.4818	0.0297	0.0257	-0.0153	0.7685	-0.6162	-0.5048				39.1770	147887
ARMA(6_4)	23	-0.0298	0.0367	-1.5488	-0.2478	0.9289	0.5493	0.0470	0.0196	1.5991	0.2987	-0.9319	-0.5458			39.1630	147890
ARMA(6_5)	15	-0.0241	0.0056	-0.2772	1.4067	0.8868	-0.6423	-0.5562	-0.0224	0.3188	-1.4263	-0.9307	0.6657	0.5970		39.0517	147837
ARMA(6_6)	11	-0.0228	0.0134	0.6885	-0.2067	0.2231	-0.3203	-0.5150	0.6510	-0.6485	0.1563	-0.1697	0.2802	0.5642	-0.6525	39.0821	147866

Table A.4.9: Output example for "CBOT Corn Futrue"

This is an estimation sample results from our program for the increments of 'CBOT Corn Futrue' data. 'Model' column is for the models we estimated, here the size of the ARMA filter is used as an indicator of the model size. 'NoS' is short for number of steps, it gives the number of steps the program needs to find the convergence of d, a minimum boundary of 9 is used for stability purpose. The value under "d, C, AR_1, AR_2, AR_3, AR_4, AR_5, AR_6, MA_1, MA_2, MA_3, MA_4, MA_5, MA_6 and Var" are models parameters for the fraction term, the constant, the AR and MA parameters and the variance of the error term, normal distribution error term is used here. The "AICBIC" is calculated by sum up AIC and BIC for model selection.

Table A.4.10: Best ARFIMA models for increments data

Product	NoS	Hurst	C	AR_1	AR_2	AR_3	AR_4	AR_5	AR_6	MA_1	MA_2	MA_3	MA_4	MA_5	MA_6	Var	AIC	BIC	Model
GSCI LEAN HOGS SPOT INDEX	45	0.3571	0.0001	1.7113	-1.5000	0.8920	-0.1077	-0.0473		-1.4371	1.2042	-0.6366				2.2036	36236	36316	ARMA(5_3)
HOGS, LEAN FUTURE CME	17	0.4141	0.0019	0.7227						-0.6232						1.6454	24642	24677	ARMA(1_1)
LB LUMBER CME	9	0.4403	0.0167	0.1137												38.2584	43082	43110	ARMA(1_0)
CBOT OATS FUTURE	9	0.4432	0.0374	-0.2377	0.7572	-0.4588	-0.9427	0.1343		0.3702	-0.7129	0.3590	0.9867			20.7434	67243	67331	ARMA(5_4)
LC CATTLE LIVE CME	11	0.4540	0.0037	0.6211	0.1967	0.6370	-0.7550			-0.5570	-0.2102	-0.6546	0.7219			0.7793	21625	21703	ARMA(4_4)
PH GOLD&SILVER INDEX	9	0.4574	-0.0069							0.0384						6.7933	37753	37781	ARMA(0_1)
NYBOT OR JUICE FUTRUE	12	0.4583	0.0030	0.7170	-0.0609	0.0408				-0.6392						5.0197	50469	50520	ARMA(3_1)
CBOT SOYBN FUTURE	9	0.4748	0.0610	1.1328	-0.9591					-1.0364	0.8646	0.0917				188.4235	92541	92600	ARMA(2_3)
CBOT WHEAT FUTURE	9	0.4775	0.0500	0.6032	-0.9352					-0.5968	0.9076					86.0859	80174	80225	ARMA(2_2)
CBOT CORN FUTURE	9	0.4815	0.0049	0.3750	0.4367	0.7561	0.0398	-0.7882		-0.3414	-0.4740	-0.7267	-0.0720	0.8131		39.0154	73854	73950	ARMA(5_5)
XAG SILVER SPOT	46	0.4870	0.0007	-0.3020	-0.4523	0.2692	0.2003	0.2935	0.5139	0.3235	0.4380	-0.1436	-0.1373	-0.2905	-0.5852	0.1261	8866	8976	ARMA(6_6)
SI SILVER NYMEX	12	0.4932	0.0014	0.8824	-0.5539	-0.6320	0.7851	-0.8875	0.0614	-0.8116	0.5398	0.6588	-0.7589	0.8369		0.1232	7560	7661	ARMA(6_5)
LME ALUMINUM SPOT	9	0.4937	-0.0303	-0.4664	0.7696	-0.3440	-0.8716			0.4432	-0.7705	0.3251	0.8663	0.0341		1171.3974	69678	69760	ARMA(4_5)
GSCI PALLADIUM INDEX ER	12	0.4972	0.1617							0.0968						59.5537	11932	11954	ARMA(0_1)
NYBOT COFFEE FUTURE	56	0.4989	0.0034	1.3817	-0.9266	0.0472	-0.0735	0.0284		-1.3832	0.9159					11.9550	56444	56517	ARMA(5_2)
CT COTTON NYBOT	38	0.5001	0.0075	-0.9929	-0.5517	-0.9342	-0.6888	0.1164		1.1121	0.6730	0.9928	0.7883			1.8064	34587	34674	ARMA(5_4)
WCE CANOLA FUTRUE	20	0.5062	0.0350	0.1940	-0.3723	0.3440	-0.6707			-0.1457	0.3200	-0.3437	0.6149			27.5791	51693	51770	ARMA(4_4)
NYBOT SUGAR FUTURE	9	0.5074	0.0008	-0.1172	0.9133	-0.2854	-0.8390	-0.0128		0.1447	-0.9186	0.2804	0.8743			0.1684	11932	12020	ARMA(5_4)
NYBOT COCOA FUTURE	99	0.5102	0.1229	1.0011	-1.1982	0.7877	-0.0893	0.0353	-0.0603	-0.9582	1.1563	-0.7196				1670.8560	114586	114674	ARMA(6_3)
GSCI COPPER INDEX SPOT	9	0.5300	0.0687	0.0293	-0.1408	0.2269	-0.6932	-0.2605	-0.3342	-0.1101	0.1227	-0.2409	0.7311	0.2105	0.2618	24.9348	58931	59039	ARMA(6_6)
CBOT SOYBN OIL FUTURE	9	0.5325	-0.0015	-0.5404	0.2072	0.4884	0.2638	-0.5212	-0.8577	0.5549	-0.2136	-0.4858	-0.2666	0.5324	0.8791	0.2249	13742	13851	ARMA(6_6)
LME 3M COPPER FUTURE	10	0.5377	0.4899	0.2071	0.0744	0.2135	-0.8684	-0.1007	-0.0585	-0.3267	-0.0764	-0.2236	0.9199			5995.0634	85512	85602	ARMA(6_4)
LME COPPER SPOT	9	0.5406	0.5708	0.1588	0.1347	0.1918	-0.8905	-0.0931	-0.0633	-0.2756	-0.1491	-0.1870	0.9305			6679.6387	86127	86217	ARMA(6_4)
HG COPPER NYMEX	9	0.5409	0.0151	0.6902	-0.6924	0.7449	-0.8107	-0.1020		-0.7955	0.7516	-0.8241	0.8809			14.8029	37005	37086	ARMA(5_4)
HO HEATING OIL NYMEX	9	0.5412	0.0356	-0.7678	-0.7566					0.7010	0.6843	-0.0947				9.4157	37076	37132	ARMA(2_3)
XPT PLATINUM SPOT	9	0.5487	0.0458	1.7174	-1.1514	0.0949				-1.6853	1.0923	-0.0567				166.5297	59189	59251	ARMA(3_3)
NYMEX CRUDE FUTURE	14	0.5512	0.0015							-0.0998	-0.0529	-0.0203	0.0239	-0.0525		1.1718	24266	24322	ARMA(0_5)
LA ALUMINUM FUTURE	18	0.5522	0.0002	0.2923						-0.4553						881.9653	43156	43188	ARMA(1_1)
PL PLATINUM NYMEX	13	0.5542	0.0259	1.6915	-1.4509	0.3930	-0.0618			-1.6617	1.3506	-0.2916				186.4083	59328	59397	ARMA(4_3)
CL CRUDE OIL NYMEX	16	0.5561	0.0033							-0.1077	-0.0560	-0.0222	0.0223	-0.0540		1.4016	21047	21102	ARMA(0_5)
LME PL INDEX	9	0.5590	0.2358	-1.6029	-0.9235					1.5754	0.8304	-0.0933	-0.0500			224.0440	48460	48520	ARMA(2_4)
WTI CUSING CRUDE SPOT	13	0.5605	0.0014							-0.1000	-0.0444	-0.0177	0.0245	-0.0691		1.2074	24284	24340	ARMA(0_5)
LA CRUDE OIL SPOT	9	0.5676	0.0012	-0.0200	0.5322	-0.0073	0.6932	0.0621	-0.7063	0.0037	-0.5634	0.0084	-0.6753	-0.0695	0.7201	0.9351	17103	17204	ARMA(6_6)
XPD PALLADIUM SPOT	9	0.5724	0.0595	0.0111	-0.0680	-0.0790										80.4283	40264	40304	ARMA(3_0)
XAU GOLD SPOT	9	0.5770	0.1570	-0.4553	0.8986	-0.2638	-0.9075	-0.0840	-0.0467	0.3969	-0.9584	0.3083	0.9100			69.1736	74450	74544	ARMA(6_4)
PA PALLADIUM NYMEX	9	0.5812	0.0991	-1.4496	-0.3201	0.3072				1.4446	0.2401	-0.4829	-0.1067			64.4031	50918	50987	ARMA(3_4)
BRENT CRUDE OIL INDEX	9	0.5948	0.0123	-1.4340	-0.8352	-0.0374	-0.1026			1.5578	0.9511					1.0513	12673	12730	ARMA(4_2)
FC CATTLE FEEDER CME	9	0.6031	0.0029	0.1132	1.2041	0.2946	-0.8585			-0.1288	-1.2402	-0.2935	0.8735			0.8173	17012	17086	ARMA(4_4)
GENERIC 1ST FUTURE GOL	11	0.6192	0.0877	0.3229	-0.4136	-0.2123	0.5577	-0.3361	-0.0751	-0.3753	0.4201	0.1839	-0.6205	0.3752		71.2587	71724	71825	ARMA(6_5)
CME FEEDER CATTLE INDEX	9	0.6560	0.0209	-0.2430	-0.2890	0.1330	0.0400	-0.1381		0.2867	0.4261					0.3847	9064	9129	ARMA(5_2)

Product	NoS	Hurst	C	AR_1	AR_2	AR_3	AR_4	AR_5	AR_6	MA_1	MA_2	MA_3	MA_4	MA_5	MA_6	Var	AIC	BIC	Model
CME FEEDER CATTLE INDEX	9	0.6680	0.0001	0.7330	0.1604	-0.1850	0.2726	0.6150	-0.6737	-0.7809	-0.1425	0.1792	-0.2484	-0.5211	0.5911	0.0000	-39517	-39420	ARMA(6_6)
GSCI PALLADIUM INDEX ER	9	0.6825	0.0003	0.0562	1.0583	-0.6937	-0.2091	0.7351		-0.1682	-1.0616	0.8483	0.1859	-0.7864	0.0460	0.0002	-9993	-9916	ARMA(5_6)
CBOT CORN FUTURE	45	0.7140	0.0001	0.8754	0.0306	0.0038	0.0322	0.0316		-0.9520						0.0001	-68405	-68339	ARMA(5_1)
PH GOLD&SILVER INDEX	99	0.7307	0.0002	0.9765						-1.0788	0.0861	0.0016	0.0380			0.0002	-43705	-43649	ARMA(1_4)
CBOT OATS FUTURE	9	0.7433	0.0002	0.7887	0.8204	-0.6075	-0.0689			-0.9216	-0.7600	0.7434				0.0002	-63720	-63647	ARMA(4_3)
WCE CANOLA FUTRUE	9	0.7566	0.0021							-0.1674						0.0001	-55903	-55875	ARMA(0_1)
HOGS, LEAN FUTURE CME	9	0.7618	0.0026							-0.2873	-0.1026	-0.0503	-0.0316			0.0003	-39114	-39066	ARMA(0_4)
NYBOT OR JUICE FUTRUE	9	0.7747	0.0004	1.4779	-0.2954	-0.2791	-0.0033	-0.0118	-0.0346	-1.6113	0.4371	0.2865				0.0002	-64489	-64401	ARMA(6_3)
WTI CUSING CRUDE SPOT	99	0.7799	0.0003	-0.5738	0.5579	0.9815				0.5478	-0.5210	-0.9011	0.0497	-0.0121		0.0003	-43292	-43215	ARMA(3_5)
NYBOT COFFEE FUTURE	9	0.7853	0.0031							-0.1441	-0.0684					0.0003	-57152	-57116	ARMA(0_2)
LC CATTLE LIVE CME	9	0.7863	0.0000	1.6689	-0.6143	-0.0233	-0.0032	-0.0294		-1.8808	0.8823					0.0001	-57430	-57360	ARMA(5_2)
FC CATTLE FEEDER CME	9	0.7878	0.0005	0.4417						-0.6705						0.0000	-47282	-47248	ARMA(1_1)
PA PALLADIUM NYMEX	9	0.8017	0.0023							-0.1565	-0.0396					0.0002	-42217	-42183	ARMA(0_2)
XPD PALLADIUM SPOT	9	0.8036	0.0030	-0.1565	-0.0630											0.0002	-31690	-31657	ARMA(2_0)
HO HEATING OIL NYMEX	9	0.8075	0.0001	1.4596	-0.3940	-0.0864				-1.6815	0.6998					0.0003	-39675	-39620	ARMA(3_2)
CBOT SOYBN OIL FUTURE	9	0.8225	0.0001	1.2391	-0.2781					-1.5202	0.5516					0.0001	-64196	-64146	ARMA(2_2)
GSCI LEAN HOGS SPOT INDEX	9	0.8248	0.0004	0.9652	-0.1812					-1.1925	0.3217					0.0001	-63866	-63816	ARMA(2_2)
CL CRUDE OIL NYMEX	9	0.8287	0.0024							-0.2615	-0.0370					0.0003	-35949	-35915	ARMA(0_2)
LA CRUDE OIL SPOT	9	0.8323	0.0061	-1.6434	-1.2267	-0.1237	0.5491	0.3867	0.0496	1.4307	0.7750	-0.3382	-0.7801	-0.3876		0.0003	-32809	-32715	ARMA(6_5)
LME PL INDEX	9	0.8324	0.0013							-0.2143	-0.0430	-0.0362				0.0001	-38396	-38356	ARMA(0_3)
XAU GOLD SPOT	10	0.8346	0.0001	0.7861	0.8102	-0.6496				-0.9233	-0.7242	0.7867	-0.0837	0.0128		0.0001	-69014	-68935	ARMA(3_5)
XPT PLATINUM SPOT	88	0.8378	0.0001	0.2724	-0.0048	-0.0071	0.0002	0.9814	-0.2716	-0.3796	0.0363	0.0179	-0.0165	-0.9231	0.3542	0.0001	-49059	-48955	ARMA(6_6)
CT COTTON NYBOT	9	0.8385	0.0008	0.4713	0.0393					-0.7452						0.0002	-57666	-57622	ARMA(2_1)
LB LUMBER CME	9	0.8436	0.0014	0.3092	1.0956	-0.6553	-0.9090	0.4610	0.0140	-0.5768	-1.1111	0.9169	0.8426	-0.7152		0.0002	-37330	-37235	ARMA(6_5)
PL PLATINUM NYMEX	9	0.8440	0.0034	-1.2567	-1.0821	-0.1917	0.3626	0.6639		0.9986	0.7383	-0.1575	-0.5153	-0.6575	0.1306	0.0001	-47428	-47331	ARMA(5_6)
CBOT SOYBN FUTURE	18	0.8460	0.0001	0.7525	0.1137	0.0134	0.0417	0.0544	-0.0051	-0.9583						0.0001	-71787	-71713	ARMA(6_1)
HG COPPER NYMEX	9	0.8473	0.0003	-0.0549	0.8837					-0.2324	-0.9791	0.2514	0.0445	-0.0040	0.0480	0.0001	-40921	-40846	ARMA(2_6)
LA ALUMINUM FUTURE	9	0.8488	0.0008	0.4041						-0.6774						0.0001	-29309	-29277	ARMA(1_1)
LME COPPER SPOT	9	0.8488	0.0002	0.0504	1.1019	0.4229	-0.7187			-0.2890	-1.1632	-0.1865	0.8611	-0.1114		0.0001	-44771	-44688	ARMA(4_5)
SI SILVER NYMEX	9	0.8535	0.0001	0.7421	0.0649	0.0425	0.0151	0.0634		-0.9441						0.0002	-57456	-57391	ARMA(5_1)
NYBOT COCOA FUTURE	9	0.8558	0.0006	0.6958	-0.0322					-1.0230	0.1735					0.0002	-64984	-64933	ARMA(2_2)
NYMEX CRUDE FUTURE	9	0.8590	0.0000	0.6262	1.1769	-0.6278	-0.1827			-0.9121	-1.0607	0.9906	0.0645	-0.0747		0.0003	-44114	-44030	ARMA(4_5)
NYBOT SUGAR FUTURE	9	0.8636	0.0012	0.4201						-0.6666						0.0003	-58643	-58606	ARMA(1_1)
XAG SILVER SPOT	9	0.8652	0.0021	-0.2309	-0.1122											0.0002	-64893	-64856	ARMA(2_0)
LME ALUMINUM SPOT	9	0.8674	0.0002	1.0289	-0.2256					-1.3163	0.4387					0.0001	-44602	-44554	ARMA(2_2)
BRENT CRUDE OIL INDEX	9	0.8693	0.0004	1.4208	-0.9879	0.2338	0.0358			-1.7298	1.3560	-0.4761				0.0001	-27026	-26962	ARMA(4_3)
CBOT WHEAT FUTURE	9	0.8703	0.0009	0.4417						-0.7054						0.0002	-64989	-64953	ARMA(1_1)
GSCI COPPER INDEX SPOT	9	0.8726	0.0001	1.3500	-0.3664	-0.0873				-1.6625	0.7314					0.0001	-60923	-60865	ARMA(3_2)
LME 3M COPPER FUTURE	9	0.8831	0.0001	1.5663	-0.6066					-1.8544	0.9855	-0.1015				0.0001	-46712	-46656	ARMA(2_3)
GENERIC 1ST FUTURE GOL	9	0.8914	0.0008							-0.2977	-0.0564	-0.0338				0.0001	-67870	-67827	ARMA(0_3)

Table A.4.11: Best ARFIMA models for log difference data

A.5 Distribution of loss

First, it is not difficult to see that for $x < 0$, we have $F_L(x) = 0$, for $x = 0$ we have $F_L(0) = P\left(I_0 \leq e^{-rT} I_T\right) = N\left(\frac{\left(\mu - r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right)$ and for $x > p_0$, we have $F_L(x) = 1$. Now let us consider $p_0 \geq x > 0$. In this case we have

$$\begin{aligned}
F_L(x) &= 1 - P(L > x) = 1 - P\left(\left(\hat{I}_T - e^{-rT} I_T\right)_+ > x\right) \\
&= 1 - P\left(\hat{I}_T - e^{-rT} I_T > x\right) \\
&= 1 - P\left(\hat{I}_T - I_0 e^{-rT} e^{\left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma B_T} > x\right) \\
&= 1 - P\left(\frac{\hat{I}_T - x}{I_0} > e^{\left(\mu - r - \frac{1}{2}\sigma^2\right)T + \sigma B_T}\right) \\
&= 1 - P\left(\frac{\log\left(\frac{\hat{I}_T - x}{I_0}\right) - \left(\mu - r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}} > B_1\right) \\
&= 1 - N\left(\frac{\log\left(\frac{\hat{I}_T - x}{I_0}\right) - \left(\mu - r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right) \\
&= N\left(\frac{\left(\mu - r - \frac{1}{2}\sigma^2\right)T - \log\left(\frac{\hat{I}_T - x}{I_0}\right)}{\sigma\sqrt{T}}\right)
\end{aligned}$$

Lower retention level is given by solving

$$N\left(\left(\lambda - \frac{1}{2}\sigma\right)\sqrt{T}\right) = F_L(0) < a = N\left(N^{-1}\left(\frac{1}{\delta}\right) + \lambda\sqrt{T}\right).$$

This gives $\left(-\frac{1}{2}\sigma\right)\sqrt{T} < N^{-1}\left(\frac{1}{\delta}\right)$ and therefore, $N\left(\left(-\frac{1}{2}\sigma\right)\sqrt{T}\right) < \frac{1}{\delta}$ and $1 - \frac{1}{\delta} < N\left(\frac{1}{2}\sigma\sqrt{T}\right)$. But since $\frac{1}{2}\sigma\sqrt{T} \geq 0$ then $N\left(\frac{1}{2}\sigma\sqrt{T}\right) \geq \frac{1}{2}$. Therefore, $\delta \leq 2$, is a sufficient condition.